VYSOKÁ ŠKOLA BÁŇSKÁ – TECHNICKÁ UNIVERZITA OSTRAVA
EKONOMICKÁ FAKULTA

KATEDRA APLIKOVANÉ INFORMATIKY

Návrh optimalizace logistických procesů v průmyslové společnosti

Design of Logistic Processes Optimization in an Industrial Company

Student: Jan Timel

Vedoucí bakalářské práce: Mgr. Ing. František Zapletal, Ph.D.

Ostrava 2016

VŠB - Technická univerzita Ostrava
Ekonomická fakulta
Katedra aplikované informatiky

Zadání bakalářské práce

Student: **Jan Timel**

Studijní program: B6209 Systémové inženýrství a informatika

Studijní obor: 6209R017 Informatika v ekonomice

Téma: Návrh optimalizace logistických procesů v průmyslové společnosti
Design of Logistic Processes Optimization in an Industrial Company

Jazyk vypracování: čeština

Zásady pro vypracování:

1. Úvod
2. Metodologická východiska v oblasti shlukové analýzy
3. Analýza vybraných logistických procesů ve zkoumaném podniku
4. Návrh a implementace optimalizačního modelu v podniku
5. Závěr

Seznam použité literatury
Seznam zkratek
Prohlášení o využití výsledků bakalářské práce
Seznam příloh
Přílohy

Seznam doporučené odborné literatury:

BUDÍKOVÁ, M., T. LERCH a Š. MIKOLÁŠ. *Základní statistické metody*. Brno: Masarykova Univerzita, 2005. ISBN 80-210-3886-1.
ŘEZANKOVÁ, H., D. HÚSEK a V. SNÁŠEL. *Shluková analýza dat*. Praha: Professional Publishing, 2007. ISBN 978-80-86946-26-9.
HAN, J., M. KAMBER and J. PEI. *Data Mining: Concepts and Techniques*. 2. ed. San Francisco: Morgan Kaufmann Publishers Inc, 2006. ISBN 1-55860-901-6.

Formální náležitosti a rozsah bakalářské práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

Vedoucí bakalářské práce: **Ing. František Zapletal, Ph.D.**

Datum zadání: 20.11.2015
Datum odevzdání: 06.05.2016

VYSOKÁ ŠKOLA BÁŇSKÁ-TECHNICKÁ UNIVERZITA OSTRAVA
Ekonomická fakulta
1

Ing. Petr Rozehnal, Ph.D.
vedoucí katedry

prof. Dr. Ing. Dana Dluhošová
děkanka fakulty

Prohlašuji, že jsem celou práci, včetně všech příloh, vypracoval samostatně.

Jan Timel

....

Jan Timel

V Ostravě dne 6. 5. 2016

Poděkování

Na tomto místě bych rád poděkoval vedoucímu mé práce Mgr. Ing. Františku Zapletalovi, Ph.D. za vedení bakalářské práce, cenné rady a připomínky. Také děkuji Mgr. Jiřímu Holubovi za poskytnutí vstupních informací důležitých k tvorbě této práce.

Obsah

1 Úvod ... 5
2 Metodologická východiska v oblasti shlukové analýzy ... 6
2.1 Stanovení podobnosti ... 7
2.1.1 Základní typy proměnných ... 7
2.1.2 Míry vzdáleností kvantitativních dat ... 8
2.1.3 Míry podobnosti nominálních dat ... 11
2.1.4 Míry podobnosti binárních dat ... 11
2.1.5 Míry podobnosti charakterizovány proměnnými různých datových typů 12
2.2 Příprava datového souboru ... 12
2.2.1 Výběr proměnných ... 12
2.2.2 Identifikace odlehlých objektů ... 13
2.2.3 Standardizace dat ... 13
2.3 Metody shlukové analýzy ... 14
2.4 Hierarchické shlukování ... 15
2.4.1 Aglomerativní přístup ... 15
2.4.2 Divizní shlukování ... 18
2.5 Nehierarchické shlukování ... 18
2.5.1 Metody s konstantním počtem shluků ... 18
2.5.2 Metody s proměnným počet shluků ... 21
2.5.3 Fuzzy shlukování ... 23
2.6 Metody pro shlukování velkých datových souborů ... 24
2.6.1 Metoda BIRCH ... 24
2.6.2 Dvoukroková shluková analýza ... 25
2.6.3 Metoda CURE ... 26
2.6.4 Metoda CLARA ... 26
2.6.5 Metoda CLARANS ... 26
2.7 Validace shlukové analýzy ... 27
3 Analýza vybraných logistických procesů ve zkoumaném podniku ... 28
3.1 Představení společnosti ... 28
3.1.1 Předmět podnikání ... 28
3.2 ČEZ Distribuční služby, s.r.o. ... 28
3.3 Odbor Správa měřidel ... 29
3.4 Popis současného stavu ... 30
4 Návrh a implementace optimalizačního modelu v podniku ... 33

4.1 Představení softwaru IBM SPSS Statistics ... 33
4.2 Stanovení parametrů optimalizace logistických procesů ... 34
4.2.1 Analýza vstupních dat ... 34
4.3 Porovnávání a výpočet vybraných metod v prostředí SPSS ... 37
4.3.1 Shlukování pomocí hierarchických metod ... 38
4.3.2 Metoda nejbližšího souseda ... 39
4.3.3 Metoda nejvzdálenějšího souseda ... 40
4.3.4 Wardova metoda ... 43
4.3.5 Ostatní hierarchické shlukovací metody ... 46
4.3.6 Metoda *k*-průměrů ... 46
4.3.7 Shlukování pomocí reálných vzdáleností ... 49
4.4 Výsledky implementace a zhodnocení výsledků ... 53
4.4.1 Stanovení nových dopravních tras ... 53
4.4.2 Ekonomické zhodnocení modelu ... 54
5 Závěr ... 56
Seznam použité literatury ... 57
Seznam použitých zkratek ... 59
Prohlášení o využití výsledků bakalářské práce
Seznam příloh
Přílohy

1 Úvod

Plánování logistických procesů je v dnešní době nedílnou součástí řízení každé firmy, která potřebuje ke své existenci přepravovat materiál nebo zboží. U organizací se lze v dnešní době setkat s vynakládáním nemalých peněžních prostředků, které slouží k zajištění optimální přepravy, a to jak z hlediska nákladů, tak i času. Zejména čas představuje v tomto procesu klíčovou roli.

Cílem bakalářské práce je analyzovat, zda je stávající řešení logistických procesů ve vybrané firmě vyhovující, nebo zda může být tento proces zdokonalován. Logistický proces je ve firmě popsán pomocí centrálních skladů a svozových míst, mezi kterými probíhá oboustranná komunikace. Výstupem práce je zhodnocení současného stavu a návrh na změnu rozložení, případně počtu skladů s ohledem na logistické náklady. Pro dosažení uvedeného cíle je použita vícerozměrná statistická metoda zvaná shluková analýza, přesněji shlukovací metody. Tyto metody spojují sledované objekty do tzv. shluků na základě jejich podobností a rozdílností. Shlukovacích metod je celá řada, jelikož se využívají v různých vědních oborech (například v biologii a informatice). V práci je popsána základní klasifikace metod pomocí skupin hierarchického a nehierarchického shlukování. Implementace ekonomického problému je v praktické části bakalářské práce řešena pomocí prostředí IBM SPSS Statistics[1].

Bakalářská práce je členěna do 5 hlavních částí. V úvodní části bakalářské práce je popsán cíl práce, uvedení do problematiky a hlavní účel vzniku práce. Druhá část se zabývá základními pojmy používanými v souvislosti se shlukovou analýzou. Jsou zde také popsány vlastnosti pro definici objektů. Následuje vysvětlení principů metod shlukové analýzy. Zvláštní pozornost je věnována metodám, které budou využity v praktické části bakalářské práce. Třetí část se zaměřuje na popis zadávající firmy a analýzu logistických procesů v této firmě. Klíčovou částí je kapitola čtvrtá, která představuje praktickou část bakalářské práce. Zde je pomocí metod shlukové analýzy provedena optimalizace a stanovení nových možností řešení dopravy v rámci modelovaného podniku. V závěrečné části bakalářské práce je provedeno shrnutí přínosů pro daný podnik a zhodnocení, zda a do jaké míry bylo dosaženo uvedených cílů.

[1] V práci je využita poslední verze IBM SPSS Statistics 23.

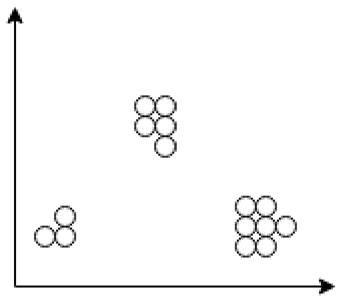
2 Metodologická východiska v oblasti shlukové analýzy

Touto statistickou metodou, anglicky Cluster analyses, se už roku 1939 zabýval Kalifornský učitel psychologie Robert. C. Tryon. Ve své publikaci tuto metodu definoval pojmem: *„Shluková analýza je obecný logický postup formulovaný jako procedura, pomocí níž seskupujeme objektivně jedince do skupin na základě jejich podobností a rozdílností.“* (Lukasová a Šarmanová, 1985. s. 15).

Dle Lukasová a Šarmanová (1985) je shlukování přirovnáváno ke klasifikaci, jakožto činnosti, vytvářející rozklad množiny objektů a vytvářející systém tříd.

„Shluková analýza slouží především jako prostředek generování hypotéz o klasifikaci objektů nebo znaků.“ (Lukasová a Šarmanová, 1985. s. 15).

Obecně je základním cílem shlukové analýzy, jakožto nástroje datové analýzy, rozdělit skupinu sledovaných objektů na co možná nejmenší počet podskupin tzv. shluků, ve které budou objekty s nejpodobnějšími vlastnostmi. Jinými slovy, aby dva objekty stejného shluku si byly více podobné než dva objekty z různých shluků. Shlukování je spjato s pojmem disjunktní množina. Dle Řezanková a kol. (2007) značí jednoznačné, nepřekrývající se přiřazení sledovaných objektů do tříd. Existují ovšem i metody, které využívají opaku disjunktnosti. Jedna z nich je uvedena v kapitole 2.4.3.



Obrázek 2.1 - Grafická prezentace disjunktních shluků, zdroj: Vlastní zpracování.

Na rozdíl od náhodného rozmístění objektů v atributovém prostoru, jsou shlukovací metody použity pro objekty, které mají tendenci se seskupovat do přirozených tříd (Horák, 2002).

2.1 Stanovení podobnosti

V literatuře se ke klasifikací objektů objevují dva přístupy. Jako první přístup je uváděno učení s učitelem. V tomto případě obsahují vstupní data, nebo datový soubor informace, o příslušnosti objektů do již známých skupin. Naopak shlukování je přiřazováno k přístupu učení bez učitele. Klade si za cíl klasifikovat všechny objekty zahrnuté do analýzy, u kterých není předem známa příslušnost žádného z objektů a ani počet skupin (Řezanková a kol., 2007).

V shlukování (zejména při využití hierarchických metod), ale i v dalších statistických metodách se vychází ze vstupní matice dat. Tato matice je tvořena řádky, které jsou představovány vektory údajů o jednotlivých objektech. Hovoříme-li o objektech, máme obvykle na mysli nějakou množinu předmětů nebo jevů. Dále je matice tvořena sloupci, které odpovídají jednotlivým proměnným (znakům nebo ukazatelům). Pro dvě kategoriální proměnné[2] je vstupem dvourozměrná tabulka četností (Řezanková a kol., 2007).

Pro potřebu zobrazovaných rovnic lze datovou matici obecně definovat jako $\mathbf{X} = m \cdot n$, s prvky x_{il}, kde $i = 1, 2, \ ..., m$ a $l = 1, 2, \ ..., n$ (Řezanková a kol., 2007).

Každý konkrétní objekt musí být tedy popsán p-ticí stavů předem stanovených p znaků. (Lukasová a Šarmanová, 1985).

Stanovení míry podobnosti se obecně liší podle toho, co je předmětem shlukování. V této kapitole jsou uvedeny jednak způsoby pro stanovení podobnosti mezi dvěma objekty, tak způsoby stanovení závislosti mezi proměnnými v objektech. V práci je detailněji popsána podobnost mezi objekty charakterizovanými kvantitativními znaky, jelikož jsou to právě ony, které jsou použity v praktické části. V kapitole 2.1.1 jsou nejprve rozčleněny typy proměnných.

2.1.1 Základní typy proměnných

Při využití shlukovacích metod je v prvé řadě definován pojem podobnosti nebo nepodobnosti objektů, přičemž je nutné rozlišit, jakými typy proměnných jsou vlastnosti jednotlivých objektů charakterizovány.

V datové matici se mohou vyskytovat proměnné nominální, ordinální, intervalové a poměrové, které lze ještě blíže specifikovat nebo zahrnout do širších skupin. Základní členění proměnných je rozdělení na kvalitativní a kvantitativní proměnné.

---

[2] Kategoriální proměnné zahrnují proměnné kvalitativní a kvantitativní diskrétní (kapitola 2.1.1).

Kvalitativní proměnné rozlišujeme:

- Nominální – tj. proměnná, o jejíž dvou hodnotách můžeme pouze říci, zda jsou stejné či různé (silnice, vozidlo).
- Ordinální – tj. proměnná, u jejíž dvou hodnot můžeme navíc určit pořadí (spokojenost).

Kvantitativní proměnné rozlišujeme:

- Intervalové – tj. proměnná, jejíž hodnoty jsou čísla, tzn., můžeme určit rozdíl dvou hodnot a mohou nabývat hodnoty 0 (teplota).
- Poměrové – tj. proměnná, jejíž hodnoty jsou pouze kladná čísla a na rozdíl od intervalových hodnot lze i určit, kolikrát je jedna hodnota větší než druhá (počet osob).

Kvantitativní proměnné mohu být diskrétní (pouze celočíselné hodnoty) nebo spojité (libovolné hodnoty z určitého intervalu).

Zvláštním typem proměnné je proměnná dichotomická, u které se při matematických výpočtech předpokládá, že jde o proměnnou binární (nabývající hodnot 0 a 1). Pro jiné statistické metody lze využít podrobnější členění. Toto členění uvádí například Budíková a kol. (2005).

2.1.2 Míry vzdáleností kvantitativních dat

V případě stanovení podobnosti resp. nepodobnosti dvojic objektů charakterizovanými kvantitativními proměnnými nemluvíme o mírách podobnosti, ale o mírách vzdálenosti.

Metriky vzdálenosti

Metriky vzdáleností vychází z geometrického modelu matice dat a prezentují objekty v prostoru. Čím je tedy vzdálenost dvou objektů menší, tím jsou si tyto objekty podobnější. Dle Řezanková a kol. (2007) se při zjišťování vzdálenosti dvou objektů x_i, x_j vychází z binární reálné funkce *d*, která splňuje následující vlastnosti:

1. $d(x_i, x_j) \geq 0$ (nezápornost) (2.1)
2. $d(x_i, x_j) = d(x_i, x_j)$ (symetrie) (2.2)
3. $d(x_i, x_j) = 0$, pro $x_i = x_j$ (identita) (2.3)

4. $d(x_i, x_j) \leq d(x_i, x_k) + d(x_j, x_k)$ (trojúhelníková nerovnost) (2.4)

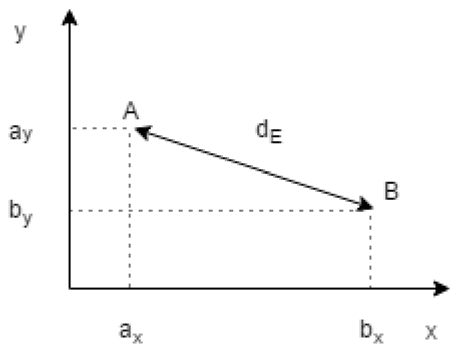
Čtvrtá podmínka (2.4) je známá jako trojúhelníková nerovnost. Požaduje tedy, aby součet velikostí odvěsen nebyl větší než délka přepony.

Euklidovská metrika

Euklidovská vzdálenost se využívá v situacích, kdy jsou proměnné nezávislé, normálně rozdělené se stejnými rozptyly. Dle Řezanková a kol. (2007) je definována vztahem (2.5).

$$d_E(x_i, x_j) = \sqrt{\sum_{l=1}^{m} (x_{il} - x_{jl})^2} \qquad (2.5)$$

Tato metrika představuje přímou vzdálenost mezi dvěma body a nebere v úvahu korelovanost proměnných (Maršálková, 2015). Graficky euklidovskou metriku znázorňuje obrázek 2.2.



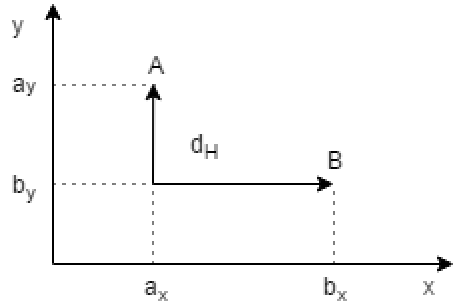
Obrázek 2.2 - Euklidovská vzdálenost, zdroj: Vlastní zpracování.

Manhattanská metrika

Za nejběžnější příklad využití této metriky vzdáleností se považuje síť New-Yorských ulic, která je pravoúhle tvořená. Ve dvourozměrném pozorování je tedy měřena po odvěsnách pravoúhlého trojúhelníku. Lze ji využít zejména pro ordinální proměnné. Je definována předpisem (2.6).

$$d_H(x_i, x_j) = \sum_{l=1}^{m} |x_{il} - x_{jl}| \qquad (2.6)$$

Graficky je Manhattanská metrika reprezentována obrázkem 2.3.



Obrázek 2.3 - Manhattanská vzdálenost, zdroj: Vlastní zpracování

Manhattanská vzdálenost aplikovaná na binární data se nazývá Hammingova (Řezanková a kol., 2007).

Čtvercová euklidovská metrika

Čtvercová euklidovská metrika využívá téměř shodného principu jako euklidovská metrika. Tvoří základ pro výpočet Wardovy metody shlukování. Dle Řezanková a kol. (2007) je stanovena vzorcem (2.7).

$$d_{ES}(x_i, x_j) = \sum_{l=1}^{m} (x_{il} - x_{jl})^2 \tag{2.7}$$

Minkovského metrika

Minkovského metrika je pouze zobecnění výše dvou uvedených metrik. Kde pro $q = 1$ jde o Hammingovu metriku a pro $q = 2$ o Euklidovu. Čím je q větší, tím více je zdůrazňován rozdíl mezi vzdálenostmi (Meloun a Militký, 2004).

Spočteme ji dle vzorce (2.8).

$$d_M(x_i, x_j) = \sqrt[q]{\sum_{l=1}^{m} |x_{il} - x_{jl}|^q} \tag{2.8}$$

Mezi další kvantitativní metriky vzdálenosti a podobnosti lze řadit dle Jarkovský a Littnerová (2011)

- průměrnou vzdálenost,
- geodetické metriky,
- Mahalanobisovu vzdálenost,

- Czekanowského koeficient,
- Whittakerův asociační index.

2.1.3 Míry podobnosti nominálních dat

Základní rozdíl mezi kvalitativními a nominálními daty je ten, že u nominálních dat lze pouze rozlišovat, zda jsou hodnoty tohoto datového typu shodné nebo různé. U těchto datových typů se používá pouze koeficientu prosté shody, který je stanoven jako podíl počtu proměnných, u nichž se hodnoty dvou objektů shodují a celkového počtu proměnných. Dle Řezanková a kol. (2007) je definován vzorcem 2.9.

$$S_{ij} = \frac{\sum_{l=1}^{m} S_{ijl}}{m}, \tag{2.9}$$

přičemž

$$S_{ijl} = 1 \ <=> x_{il} = x_{jl},$$

$$S_{ijl} = 0 \text{ v ostatních případech,}$$

kde S_{ijl}vyjadřuje podobnost mezi *i*-tým a *j*-tým objektem na základě *l*-té proměnné.

2.1.4 Míry podobnosti binárních dat

V případě binárních dat se v mnoha literaturách (například v Řezanková a kol. (2007)) hovoří o tzv. asociačních mírách. Používají se pro porovnání objektů, jejichž proměnné jsou nemetrického[3] charakteru.

Tento typ proměnných je rovněž dle Řezanková a kol. (2007) velmi často využíván i v dalších vědních disciplínách jako je medicína, informační technologie aj. Rovněž proměnné kteréhokoliv z dalších datových typů, lze převést na hodnoty 0 a 1 (například ekonomicky aktivní a neaktivní). Za nejběžnější členění měr binárních dat považujeme členění na míry pro symetrické a asymetrické proměnné.

V případě symetrických proměnných jsou využívány Sokalův-Michenerův a Rogersův-Tanimotoův koeficient. A v případě asymetrických proměnných lze uvést Jacardův, Diceův a

[3] Proměnné nesplňující čtyři základní vlastnosti metrického prostoru uvedeného v kapitole 2.1.2.

Czekanowského koeficient. Koeficienty jsou podrobněji rozebrány v Řezanková a kol. (2007) nebo Meloun a Militký (2004).

2.1.5 Míry podobnosti charakterizovány proměnnými různých datových typů

V případě proměnných různých datových typů je dle Řezanková a kol. (2007) pro stanovení míry podobnosti uváděn Gowerův koeficient podobnosti.

2.2 Příprava datového souboru

Po vytvoření datové matice by měla být věnována pozornost předzpracování dat. Pokud chceme dosáhnout co nejspolehlivějších a nejúplnějších výsledků prostřednictvím metod shlukové analýzy, musí datový soubor projít úpravou. Dle Meloun a Militký (2004) úprava rovněž ovlivňuje i metriky vzdálenosti, jelikož většina z nich je velmi citlivá na měřítka vedoucí k různé numerické velikosti proměnných. Jednotlivé kroky tohoto postupu jsou

- výběr proměnných,
- standardizace dat,
- identifikace odlehlých objektů.

2.2.1 Výběr proměnných

V ideálním případě by dle Řezanková a kol. (2007) bylo vhodné a některé metody to i vyžadují, aby v souboru zůstaly pouze proměnné statisticky nezávislé. Neexistuje ale univerzální ukazatel statistické závislosti, neboť pro každou dvojici proměnných, obsažených v datové matici se využívá odlišný přístup.

Vzhledem k tomu, že se tato práce zabývá výhradně datovými soubory obsahující proměnné kvantitativní, je pro tento typ proměnných vhodným nástrojem posouzení statistické závislosti Pearsonův korelační koeficient (2.10).

$$r_{kl} = \frac{\sum_{i=1}^{n}(x_{ik}-\bar{x}_k)(x_{il}-\bar{x}_l)}{\sqrt{\sum_{i=1}^{n}(x_{ik}-\bar{x}_k)^2\ \sum_{i=1}^{n}(x_{il}-\bar{x}_l)^2}}, \tag{2.10}$$

kde $\bar{x}_k$ je aritmetický průměr hodnot k-té proměnné a $\bar{x}_l$ aritmetický průměr l-té proměnné (Řezanková a kol., 2007).

Výsledky koeficientu korelace nabývají hodnot z intervalu $\langle -1,1\rangle$. Pokud je výsledná hodnota blízká některému z krajních bodů, lze tyto proměnné považovat za závislé. Záporná závislost je pozorována v případě záporného výsledku koeficientu korelace, při kladných výsledcích tohoto koeficientu se objevuje závislost kladná.

Hodnota korelačního koeficientu -1 značí zcela nepřímou závislost, mnohdy taky uváděnou jako antikorelaci. Obecně značí, že čím více se zvětší hodnoty v první skupině znaků, tím více se zmenší hodnoty v druhé skupině znaků, např. vztah mezi uplynulým a zbývajícím časem. Pokud je hodnota korelačního koeficientu rovna 1, značí zcela přímou závislost, např. vztah mezi rychlostí bicyklu a frekvencí otáček kola bicyklu. Naopak proměnné lze považovat za nezávislé, pokud hodnota dvou srovnávaných proměnných je blízká nule.

2.2.2 Identifikace odlehlých objektů

Shlukovací metody jsou také citlivé na odlehlá pozorování. Výsledná kvalita shluku, může být odlehlými hodnotami negativně ovlivněna. Měla by být tudíž věnována velká pozornost, zda do analýzy budou zahrnuty všechny proměnné, nebo jen některé. Pomocí vhodné přípravné fáze lze tyto pozorování odhalit. Druhou možností je způsob provedení několika kroků shlukovacích algoritmů, ve kterých se mohou oddělit shluky s malým počtem objektů (například pomocí dvoufázového algoritmu *k*-průměrů (2.5.1)). Druhý zmiňovaný způsob ale není použitelný pro velké datové soubory z důvodu velké časové náročnosti jednotlivých kroků (Žambochová, 2010).

Výstupem obou těchto postupů je identifikace odlehlých objektů, jejich eliminace z matice vstupních dat a následné nezahrnutí do shlukovacích algoritmů.

2.2.3 Standardizace dat

Posledním krokem je standardizace dat (rovněž také uváděna jako transformace dat). Rozumí se jí odstranění závislosti na jednotkách znaků a rozptýlení. V datovém souboru se lze často setkat s proměnnými, které se liší v jednotkách. Vyskytují se zde například hodnoty naměřené v gramech, ve stupních, v metrech. Do značné míry poté může jeden druh proměnné převažovat nad ostatními a přitom nežádoucím způsobem ovlivňovat výsledek pozorování. Úkolem standardizace dat je tedy zajistit, aby každý znak ovlivnil průběh shlukování stejnou měrou.

Standardizace může být dvojího typu. Nejčastěji se používá standardizace sloupcová, která v matici dat normalizuje proměnné do svého *z*-skóre, tj. prvky původních dat *l*-tého znaku x_{il} jsou po odečtení aritmetického průměru děleny svou směrodatnou odchylkou s_l.

$$z_{il} = \frac{x_{il} - \bar{x}_l}{s_l}, \tag{2.11}$$

přičemž

$$\bar{x}_l = \sum_{i=1}^{n} \frac{x_{il}}{n} \tag{2.12}$$

a

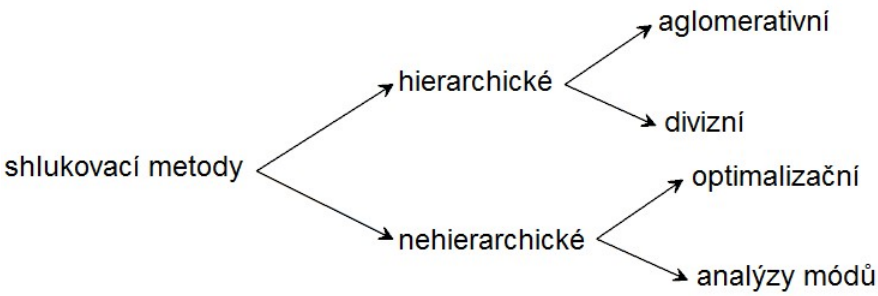
$$s_l = \sqrt{\frac{\sum_{i=1}^{n}(x_{il} - \bar{x}_l)^2}{n-1}}. \tag{2.13}$$

Takto standardizované proměnné mají střední hodnotu rovnou 0 a rozptyl 1 (Řezanková a kol., 2007).

Druhým typem je standardizace řádková, která ale není vhodná k identifikací shluků dle vzdáleností. Může být však efektivní ve speciálních případech např. pro odstranění nekonstantnosti rozptylu (Rybová, 2015).

2.3 Metody shlukové analýzy

Obecně se dnes v literatuře vyskytuje velké množství metod, které není lehké jednoznačně zařadit do některé z kategorií. Většina kategorií se může překrývat, takže metoda může obsahovat rysy hned několika z nich. Dle Lukasová a Šarmanová (1985) lze shlukovací metody dělit podle schématu uvedeného v obrázku 2.4.



Obrázek 2.4 - Základní skupiny metod shlukové analýzy (Lukasová a Šarmanová, 1985).

2.4 Hierarchické shlukování

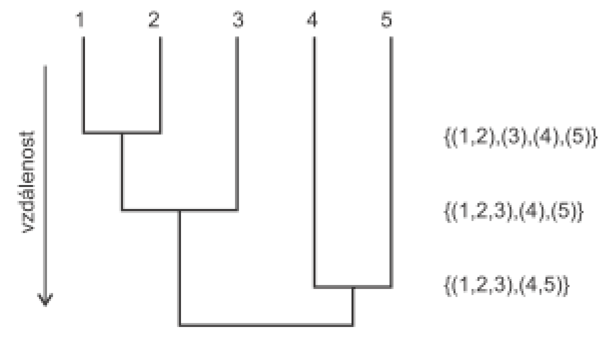
Hierarchická struktura shlukování může být zapříčiněna dvěma základními způsoby. Tyto způsoby jsou nazývány aglomerativní shlukování a divizní shlukování. Přičemž základní vlastností hierarchických metod shlukování je skutečnost, že jednotlivé kroky na sebe navazují. Výsledky předešlého kroku jsou vždy přiřazeny k získaným výsledkům v následujícím kroku, a je tak vytvářena stromová struktura.

2.4.1 Aglomerativní přístup

Na počátku tvoří každý objekt určité množiny jednoprvkový shluk. Tento krok definujme jako nultý rozklad Ω_0 množiny objektů. V následujícím kroku se zvolí v množině objektů dva, které jsou si dle daného kritéria nejvíce podobné a spojí se do nově vzniklého shluku Ω_1. Daným kritériem se rozumí výběr vhodné shlukovací metody. Tento proces se opakuje až do doby, než dospějeme k poslednímu rozkladu Ω_{n-1} o jediném shluku zastřešující všechny ostatní shlukované objekty (Lukasová a Šarmanová, 1985).

Dendrogram

Dendrogramem se rozumí grafické zobrazení metod hierarchického shlukování. Jde o stromový diagram, který v první fází zobrazí jednotlivé objekty odděleně. Následně krok za krokem spojuje objekty do nových shluků podle kritéria nejmenší vzdálenosti. V posledním kroku vytvoří jeden shluk, který zastřešuje všechny ostatní. Dendrogram je rovněž empirickou pomůckou pro nalezení výsledného počtu shluků v hierarchických metodách (obrázek 2.5).



Obrázek 2.4 - Dendrogram pěti objektů (Kelbel a Šilhán, 2002).

Jako další graf znázorňující hierarchické shlukování lze uvést rampouchový graf (anglicky icicle plot) nebo banner plot, což je jeho horizontální obdoba. Podrobnější popis těchto grafů lze nalézt v Řezanková a kol. (2007).

Metoda nejbližšího souseda

Postup metody nejbližšího je sestaven na minimálních vzdálenostech mezi jednotlivými objekty. Stanoví se dva shluky, které spojuje nejkratší vzdálenost v datovém souboru a vytvoří se z nich shluk nový. Následuje přidání třetího objektu do shluku, rovněž podle minimální vzdálenosti. Proces končí přidáním posledního objektu do nového shluku. Ve výsledném dendrogramu (2.5) lze vidět postupné řetězení shluků. Velkou nevýhodu této metody lze považovat v tom, že ve výsledném shluku lze nalézt objekty nebo shluky, které by bylo možné jinak oddělit do samostatných shluků.

Metoda nejvzdálenějšího souseda

Princip metody nejvzdálenějšího souseda je velice podobný metodě nejbližšího souseda. Objekty se rovněž spojují do shluků podle nejkratších vzdáleností, ale tato vzdálenost je dána největší vzdáleností od jeho členů tzn. nejvzdálenějších objektů mezi posuzovanými shluky. Rozdíl je tedy v tom, že objekty ve shluku mohou být k sobě už jen blíže nebo stejně, než je tato vzdálenost. Oproti metodě nejbližšího souseda zabraňuje řetězení shluků a je vhodná zejména pro shluky o stejné velikosti (Jarkovský a Littnerová, 2011).

Metoda průměrné vazby

V metodě průměrné vazby se vzdálenost mezi m-tým shlukem a shlukem n-tým rovná průměrné vzdálenosti kteréhokoliv člena m-tého shluku od kteréhokoliv člena n-tého shluku. Nejvíce podobné jsou si tedy ty shluky, jejichž průměrná vzdálenost je nejmenší. Dle Han a kol. (2006) je tato metoda oproti dvěma výše uvedeným metodám méně citlivá na odlehlá pozorování díky průměrné hodnotě. Rovněž díky jednoduchému výpočtu průměrné vzdálenosti je tato metoda vhodná k zpracování kategorických i numerických dat.

V programové příručce SPSS i v literatuře (např. Řezanková a kol. (2007)), lze nalézt detailnější dělení této metody. Kromě výše zmíněného principu lze tuto metodu rozdělit na metodu průměrné vazby pro vnitroshlukové vzdálenosti. Rozdíl spočívá v tom, že nejprve se objekty dvou shluků spojí do jednoho shluku a následně se vypočte aritmetický průměr vzdáleností.

Centroidní metoda

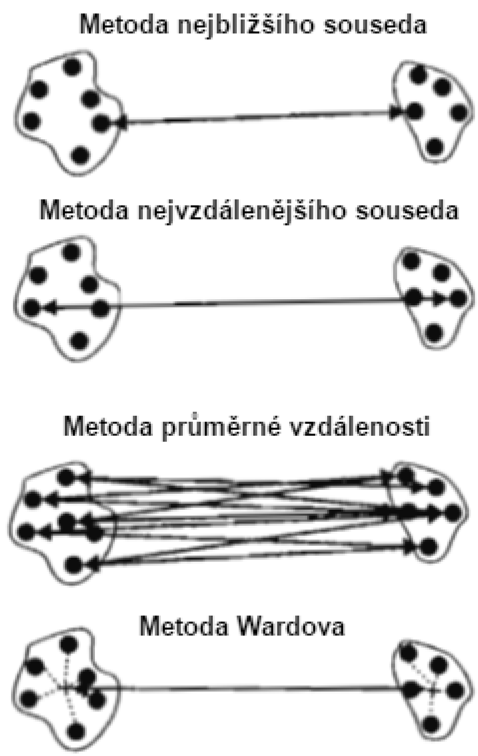
Pro zjištění nepodobnosti shluků je zde využito euklidovské vzdálenosti, přesněji čtverce euklidovské vzdálenosti d_E^2. Dále je zde uveden pojem centroid. Centroidem, neboli těžištěm, se dle Řezanková a kol. (2007) rozumí vektory aritmetických průměrů, vzniklé na základě všech objektů v daném shluku. Shluky se tedy spojují podle nejmenší čtvercové euklidovské vzdálenosti těžišť, která se po každém kroku znovu přepočítává.

Na obdobném principu je založena mediánová metoda. Oproti centroidní metodě přiřazuje jednotlivým shlukům podle jejich velikosti rozdílné váhy.

Wardova metoda

Wardova metoda je založena na minimalizaci ztráty informací při vytváření nových shluků. Metoda se tedy nezaobírá vzdáleností mezi shluky. Při vytvoření nového shluku se pro všechny dvojice odchylek vypočítá přírůstek sumy čtverců odchylek, které vznikly jejich sloučením. Výsledné shluky vzniknou na základě spojení shluků s minimální hodnotou tohoto přírůstku. Shluky stanovené Wardovou metodou mají relativně stejné velikosti (Meloun a Militký, 2004).

Grafické znázornění aglomerativních metod dle Meloun a Militký (2004) je zobrazeno v obrázku 2.6.



Obrázek 2.5 - Nejčastěji používané aglomerativní metody (Meloun a Militký, 2004).

2.4.2 Divizní shlukování

Divizní shlukování je do jisté míry opakem aglomerativního přístupu. Za počáteční stav lze zde považovat množinu objektů, která je tvořena pouze jedním shlukem. Následně probíhá rozdělení na menší a menší shluky, až zůstanou pouze shluky jednoprvkové. Rozklady shluků na menší části jsou rovněž realizované podle určitého kritéria. Vzhledem k velkému množství způsobů rozkladu je tato možnost aplikovatelná spíše pro velmi malý počet objektů (Lukasová a Šarmanová, 1985).

Jako jeden z mnoha příkladů je v literatuře uváděna MacNaughton-Smithova metoda. Podrobněji je tato metoda rozebrána v Lukasová a Šarmanová (1985).

2.5 Nehierarchické shlukování

Dle Řezanková a kol. (2007) lze konstatovat, že na klasifikaci metod nehierarchického shlukování neboli metod rozkladu, existuje velké množství pohledů. Autoři zde zařazují iterativní relokační algoritmy, metody matematického programování, metody založené na hustotě a dále metody využívající převážně kvantitativné proměnné.

Obecně lze uvést, že nehierarchické shlukování se oproti hierarchickému shlukování liší především stanovením počátečních shluků. Dále se nehierarchické shlukování rozlišuje na metody, u kterých se stanovený počet shluků během provádění shlukovacích metod mění (tyto metody lze nazvat metodami s optimalizačním neboli proměnným počtem shluků) nebo zůstává konstantní (metody s konstantním počtem shluků).

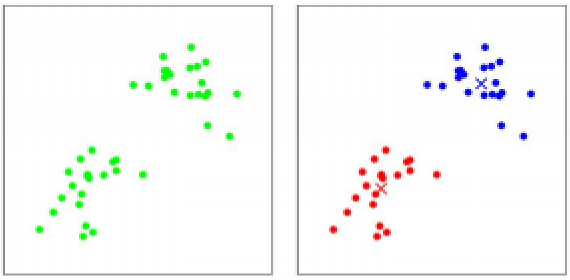
Meloun a Militký (2004) uvádí navíc členění na metody označovány jako analýza módů, které představují hledání rozkladu do shluků. Tyto shluky jsou charakterizovány zvýšenou koncentrací objektů v *m*-rozměrném prostoru znaků.

2.5.1 Metody s konstantním počtem shluků

Metoda *k*-průměrů

Metoda *k*-průměrů pracuje pouze s daty, které jsou charakterizovány kvantitativními proměnnými a díky absenci matice vzdálenosti je vhodná i pro velké datové soubory. Proces této metody je založen na systematickém přesunu objektů mezi nově vznikajícími shluky. V prvním kroku metody dochází k výběru *k* počátečních bodů. Dle Řezanková a kol. (2007) jsou tyto body nazývány centroidy. V jiných literaturách, například v Meloun a Militký (2004), se lze setkat s výstižnějším označením jako shlukováním metodou nejbližších těžišť. Centroidem se v případě metody *k*-průměrů rozumí vektor, pro který platí, že součet euklidovských vzdáleností je k tomuto vektoru minimální. Centroid se stanovuje pro každý nově vznikající shluk pouze jeden. Výběr těchto počátečních centroidů lze provést pomocí různých přístupů. MacQueen, který je dle Lukasová a Šarmanová (1985) autorem této metody, doporučuje vybrat prvních *k* objektů z posuzovaného datového souboru. Metoda pokračuje druhým krokem, ve kterém se jednotlivé objekty datového souboru přiřazují ke zvoleným centroidům podle minimální euklidovské vzdálenosti. Po vyčerpání a přiřazení všech objektů mezi jednotlivé centroidy, přičemž dochází ke vzniku shluků nových, následuje opět přepočet vzdálenosti k jednotlivým centroidům. Pokud má objekt blíže k centroidu shluku jiného, proběhne přesun objektu do nového shluku. Celý tento proces končí ve chvíli, kdy se mezi objekty nemění příslušnost k daným shlukům. První a poslední fázi metody *k*-průměrů zobrazuje obrázek 2.7.

Dle Košťál (2013) mohou být výsledky této metody ovlivněny pořadím pozorování v souboru. Proto je tedy nutné jednotlivé kroky metody opakovat (například s využitím náhodné rotace a s rozdělením souboru na polovinu). Postupnými kroky by se rovněž měly vylučovat odlehlá pozorování.



Obrázek 2.6 - První a poslední fáze algoritmu *k*-průměrů. Upraveno dle Piech (2013).

Základní vlastnosti metody *k*-průměrů lze shrnout dle Kelbel a Šilhán (2002) do těchto 5 bodů:

- metoda je jednoduchá,
- během algoritmu dochází ke změnám příslušnosti objektů ve shlucích,
- po konečném počtu kroků konverguje k určitému řešení,
- v závislostech na počátečních podmínkách lze předpokládat možnost více řešení,
- metoda je citlivá na odlehlá pozorování.

Úpravou metody k-průměrů vznikly další metody zachovávající stanovený počet shluků.

Metoda *k*-medoidů

Metoda bývá rovněž označována zkratkou PAM (Partitioning Around Medoids). Podstata metody je velice podobná metodě k-průměrů. Vychází totiž také z počátečního rozložení sledovaných objektů do k-shluků. V první fázi této metody se pro každý vytvořený shluk stanoví medoid. Medoidem se rozumí objekt ze shluku, ke kterému je součet vzdáleností ostatních objektů v daném shluku minimální. Následně se zkoumají ostatní objekty a probíhá přiřazování podle vzdálenosti k medoidům z různých shluků. Objekty tedy mohou zůstat ve shluku s prvotním medoidem nebo budou přesunuty do shluku k medoidu, ke kterému mají blíže (Řezanková a kol., 2007). Objekt x_i je tedy umístěn do shluku C_g pokud platí

$$d(x_i, m_g) \leq d(x_i, m_u), \tag{2.14}$$

přičemž m_g a m_u značí rozdílné medoidy pro všechna $u = 1,2, \dots, k$ (Řezanková a kol., 2007).

V dalších krocích se minimalizuje hodnota funkce (2.15), která pro každý shluk odpovídá součtu vzdálenosti všech objektů od medoidu v daném shluku.

$$f = \sum_{i=1}^{n} d(x_i, m_{g,i}), \tag{2.15}$$

kde $m_{g,i}$ představuje medoid v g-tém shluku, k němuž je přiřazen i-tý objekt. Proces je ukončen ve chvíli, kdy přestane klesat hodnota funkce (2.15).

Oproti metodě k-průměrů nelze metodu PAM zkreslit za pomoci odlehlých pozorování. Na druhou stranu ale není vhodná pro větší datové soubory.

Metody *k*-módů a *k*-histogramů

Jedná se o modifikace metody k-průměrů, které jsou určeny pro shlukování objektů charakterizovaných pomocí nominálních proměnných. V případě algoritmu k-modů se používá koeficient prosté shody (2.9), přesněji je z tohoto koeficientu odvozena míra nepodobnosti. Dále se je těmto metodám věnováno například v Řezanková a kol. (2007).

2.5.2 Metody s proměnným počet shluků

Zásadní rozdíl u metod měnicí počet shluků je ten, že oproti metodám uvedených v kapitole 2.5.1 umožňují slučování a rozdělování shluků objektů během shlukovacího procesu. U všech metod se rovněž nastavuje počáteční počet *k* shluků. Výsledný počet shluků tedy nemusí odpovídat počtu shluků stanovených na začátku. Dále je pro ně charakteristické stanovení dalších počátečních (dále uváděny jako řídící) parametrů, u kterých je rozhodnuto, zda se změny v počtu skupin budou provádět či nikoliv (Kelbel a Šilhán, 2002).

Tyto řídící parametry se dle Lukasová a Šarmanová (1985) dělí obecně do dvou typů:

1. Řídící parametry jsou zadány analytikem a zůstávají během celého procesu konstantní. Zde jsou zařazeny MacQueenova metoda se dvěma parametry, Wishartova metoda RELOC a metoda ISODATA.
2. Řídící parametry jsou stanoveny z analyzovaných dat. V této skupině se nachází metoda CLASS.

V rámci krátkého popisu jednotlivých algoritmů se zaměřuji zejména na popis vstupních parametrů.

MacQueenova metoda se dvěma parametry

Algoritmus této metody je složen ze dvou částí. První z nich je v Lukasová a Šarmanová (1985) pojmenována jako algoritmus TYPM a druhou částí je algoritmus MACQ, který je identický s MacQueenovou metodou *k*-průměrů.

Na začátku procedury TYPM se stanovují tři vstupní parametry:

- počet centroidů *k*,
- slučovací parametr *C*,
- rozdělovací parametr *R*.

Dle Lukasová a Šarmanová (1985) se provede dosazení prvních *k* bodů za centroidy a následují dva kroky:

1. Mezi jednotlivými centroidy se stanoví vzdálenost. Pokud je nejmenší z těchto vzdáleností menší než *C*, následuje sjednocení skupin bodů s nejmenší vzdáleností centroidů a umístění nově vzniklého centroidu do těžiště vzniklé skupiny. Tato

iterace končí v okamžiku, kdy jsou všechny centroidy navzájem vzdáleny alespoň C.

2. V druhém kroku se provádí postupné přiřazení zbývajících bodů k existujícím centroidům. Jestliže je vzdálenost bodu od nejbližšího centroidu maximálně R, stává se součástí nejbližšího centroidu a daný centroid se přesune do těžiště zvětšené skupiny bodů. Následně se provádí znovu kontrola existujících centroidů. Je-li ovšem vzdálenost bodu větší než R, stává se novým centroidem.

Na výsledné centroidy se posléze aplikuje metoda k-průměrů.

Wishartova metoda RELOC

Metoda RELOC vyžaduje zadání čtyř řídících parametrů: vzdálenostní práh, minimální počet objektů ve shluku, minimální počet shluků a maximální počet jednotlivých iterací.

V první fázi je zapotřebí provést počáteční rozklad bodů do k shluků. Dle Lukasová a Šarmanová (1985) MacQueen navrhuje vybrat k počátečních bodů[4] z libovolně uspořádané množiny, přičemž počet shluků k musí být větší než stanovený minimální počet shluků. Následně probíhá určení centroidů v každém shluku. Proces celého algoritmu zahrnuje střídavé provádění vytváření a rozpuštění shluků do té doby, kdy budou prováděny přesuny jednotlivých bodů k jiným centroidům nebo počet iterací dosáhne stanovené hodnoty. Postup vytváření a rozpouštění shluků je popsán v Hynar (2003).

Metoda ISODATA

ISODATA má podstatně více parametrů než předchozí dvě metody. Kromě již zmíněného počtu vstupních shluků k je nutno před aplikací algoritmu zadat: povolenou vnitroshlukovou směrodatnou odchylku, minimální počet objektů ve shluku, minimální vzdálenost mezi dvěma centry[5] shluků, očekávaný počet shluků, maximální počet dvojic shluků, které se spojují v rámci jedné iterace, a maximální počet iterací. Jednotlivé kroky algoritmu jsou popsány v Žambochová (2010) nebo Lukasová a Šarmanová (1985).

[4] V tomto případě se jedná o objekty stanovené analytikem, kolem nichž se dá očekávat vytvoření shluků.
[5] Centrem lze stanovit centroid nebo medoid.

Metoda CLASS

Vznikla jako modifikace metody ISODATA a jak už bylo uvedeno na začátku kapitoly 2.5.2, většina parametrů je stanovena automaticky, tudíž je metoda méně závislá na schopnostech analytika. Parametry jsou zde pouze tři. Prvním z nich je maximální počet iterací, minimální počet objektů ve shluku a počáteční rozdělovací práh. Jednotlivé kroky algoritmu zahrnují vyloučení malých skupin, rozdělení skupin a zrušení skupiny (Lukasová a Šarmanová, 1985).

Stejně jako ve výše zmiňovaných postupech, metoda končí optimálním přiřazením objektů do shluků nebo dosažení maximálního počtu iterací.

2.5.3 Fuzzy shlukování

Fuzzy analýzy lze řadit do kategorie nehierarchického shlukování, i když ve své podstatě se od ostatních metod liší. Doposud zmiňované metody jsou založeny na principu jasného přiřazení objektu do shluku. Ve fuzzy analýze jsou objekty přítomné ve všech shlucích. Fuzzy tedy reprezentuje princip pro zachycení neurčitosti. V prvé řadě je přiřazena každému objektu míra příslušnosti ke každému ze shluků a posléze provedeno binární přiřazení. V definované míře příslušnosti u_{ih} musí platit $0 \leq u_{ih} \leq 1$ a suma těchto hodnot u_{ih} musí být rovna 1.

Výhodou fuzzy shlukování je vícenásobná příslušnost, ovšem vyskytuje se zde mnohem více informací, které musí být popsány. Dle Řezanková a kol. (2007) lze algoritmus definovat:

1. Fuzzy algoritmus minimalizuje účelovou funkci f

$$f = \sum_{h=1}^{k} \frac{\sum_{i=l}^{n} \sum_{j=1}^{n} u_{ih}^2 u_{jh}^2 d_{ij}}{2 \sum_{j=1}^{n} u_{jh}^2}, \tag{2.16}$$

kde u_{ih} představuje neznámou míru příslušnosti objektu i v h-tém shluku a d_{ij} je vzdálenost mezi objekty *i* a *j*.

Ohodnocení fuzzy shlukování stanovuje Dunnův koeficient rozkladu (2.17), který představuje míru, jak těsně padne fuzzy řešení na příslušné pevné shluky. Za pevné shluky lze dle Meloun a Militký (2004) považovat klasifikaci každého objektu do shluku, v němž má největší u_{ih}.

$$F = \frac{1}{n}\sum_{i=1}^{n}\sum_{h=1}^{k} u_{ih}^{2}\,, \tag{2.17}$$

přičemž koeficient nabývá hodnot intervalu $\langle\frac{1}{k}; 1\rangle$. Hodnoty $\frac{1}{k}$ lze dosáhnout, pokud se všechna $u_{ih} = 1$, pro $i = 1, \dots, n$ a $h = 1, \dots, k$. Tato situace se tedy nazývá úplné fuzzy shlukování. Pro pevné shlukování platí, že právě jedno $u_{ih} = 1$ přičemž všechna ostatní $u_{ih} = 0$.

Problematika fuzzy analýzy je ovšem rozsáhlejší. V Meloun a Militký (2004) se lze setkat s dalším koeficientem, jmenovitě Kaufmanovým rozdělovacím koeficientem (2.18).

$$D = \frac{1}{n}\sum_{i=1}^{n}\sum_{h=1}^{k} (g_{ih} - u_{ih})^{2}\,, \tag{2.18}$$

kde g_{ih} a u_{ih} značí pravděpodobnosti přiřazení objektu *i* buďto do shluku *k* nebo *j*.

2.6 Metody pro shlukování velkých datových souborů

Hierarchické i nehierarchické shlukování tvoří základ pro celou řadou metod, které se využívají ve vícefázovém shlukování. Tyto metody jsou navrženy zejména pro velké datové soubory. V literatuře je popsáno několik těchto specifických metod (přesněji algoritmů) (Řezanková a kol., 2007).

Pro hierarchický přístup jsou v práci uvedeny algoritmy:

- BIRCH (Balanced iterative reducing and clustering using hierarchies),
- CURE (Clustering using representatives).

Základem těchto metod je postup, který se v případě hierarchického shlukování nazývá frakcionalizace. V první fázi se datový soubor rozděluje do tzv. frakcí neboli podsouborů, přičemž je v každé frakci stanoven určitý počet shluků. Pro každou vybranou frakci se v dalším kroku spočte centroid. Takto vzniklé shluky jsou dále shlukovány do *k* předem stanovených skupin s využitím stejné metody shlukové analýzy. Objekty, které zůstaly nepřiřazené, se do vytvořených shluků přiřazují na základě centroidů (Řezanková a kol., 2007).

2.6.1 Metoda BIRCH

Díky kombinaci hierarchických a jiných metod, zejména opakujících se rozdělení, překonává metoda BIRCH dle Han a kol. (2006) dva nedostatky aglomerativních metod.

Prvním z nich je škálovatelnost a druhým velká závislost na předchozím kroku. V této metodě jsou zavedeny dva pojmy: shlukovací vlastnosti, označované zkratkou CF a shlukovací strom[6] (CF strom), který slouží k vytváření a uchovávání shlukovacích vlastností. CF je tříprvkový vektor obsahující:

- počet objektů v daném podshluku (n),
- vektor, jakožto součet hodnot příslušné proměnné (LS),
- další vektor, jehož každý prvek se rovná součtu druhých mocnin příslušné proměnné (SS).

Matematické vyjádření odpovídá:

$$\boldsymbol{CF} = (n, LS, SS) \tag{2.19}$$

Úvodní charakteristika metody BIRCH je zde uvedena zejména proto, že ve statistickém programovém prostředí SPSS je její princip využit pro dvoukrokovou shlukovou analýzu[7]. Problematika metody BIRCH je dále rozebírána například v Zhang a kol. (1996).

2.6.2 Dvoukroková shluková analýza

V této proceduře lze pracovat s kvantitativní i kvalitativními proměnnými. Metoda dvoukrokové analýzy se obecně skládá ze dvou částí. V první části se objekty shlukují do podshluků s využitím inkrementálního shlukování. Inkrementální shlukování hodnotí objekty buďto přiřazením do již vytvořeného shluku nebo vytvořením shluku nového (tedy jednoprvkového). Zde jsou využity poznatky, které jsou popsány v metodě BIRCH. Jednotlivé podshluky tedy tvoří listy CF stromu. Pro každý vstup do jednotlivých uzlů stromu se stanovuje CF charakteristika, kterou jsou charakterizovány všechny objekty daného vstupu. Objekty vstupují do jednotlivých podshluků na základě vzdálenosti. Při každém novém vstupu je CF charakteristika aktualizována. Ve druhé fázi jsou všechny podshluky zařazeny do předem stanoveného počtu shluků. Na základě poznatku, díky kterého je výsledný počet shluků menší než počet objektů původního souboru, mohou být dále aplikovány tradiční metody hierarchického shlukování. V obou fázích je využita shodná míra nepodobnosti. Přesněji euklidovská vzdálenost, kterou lze využít pro kvantitativní spojité proměnné a míra

[6] Tvoří jakési schéma, jehož prvky jsou tvořeny uzly a větvemi.
[7] V prostředí SPSS se jedná o proceduru TwoStep Cluster analyses.

nepodobnosti typu věrohodnostního poměru určená pro kategoriální proměnné (Řezanková a kol., 2007).

2.6.3 Metoda CURE

Algoritmus CURE je velice vhodný pro datové soubory, které obsahují ve svých datech shluky různých tvarů a značné odchylky. Velká pozornost je zde zaměřována na centroidní body a odlehlá pozorování. Pracuje se zde pouze s náhodně vybranými objekty z předem připravených frakcí. Podstata této metody je, že pro každý shluk je vybrán konstantní počet *c* reprezentantů. V prvním kroku je z datového souboru vybrán náhodný vzorek objektů, které už jsou rozděleny do zmiňovaných frakcí. Dále algoritmus využívá metod hierarchického shlukování pro spojení objektů do pomocných shluků a identifikace odlehlých objektů na základě minimální vzdálenosti dvou objektů. Po vyloučení odlehlých objektů je vybrán konstantní počet *c* reprezentantů. V poslední fázi se dosud nezahrnuté objekty přiřadí do shluku, jehož reprezentant je nejblíže k danému objektu (Řezanková a kol., 2007).

Pro nehierarchický přístup jsou v literatuře nejčastěji uváděny metody CLARA a CLARANS, které vycházejí z metody PAM (kapitola 2.5.1).

2.6.4 Metoda CLARA

Metoda *k*-průměrů je modifikována do tří, několikrát se opakujících kroků. V prvním kroku se provede náhodný výběr předzpracovaných objektů, které jsou rozděleny do *k* skupin. Zbylé objekty se ve druhém kroku přiřadí do vzniklých skupin na základě nejkratší vzdálenosti. V posledním kroku je zapamatována průměrná vnitroskupinová vzdálenost. Po opakovaných pokusech je zvoleno shlukování na základě nejmenší průměrné vzdálenosti (Řezanková a kol., 2007).

Mezi hlavní slabinu metody CLARA lze řadit vysokou závislost na vybraném počtu objektů (Zaiane, 1999).

2.6.5 Metoda CLARANS

Metoda CLARANS je velice podobná principu metody CLARA. Dle Han a kol. (2006) se samotný proces dá popsat pomocí prohledávání stromového grafu, ve kterém může být každý uzel tím správným medoidem. Jednotlivé uzly spolu sousedí, pokud mají dva rozdílné medoidy.

V prvním kroku se vstupním parametrem stanovuje počet sousedů zvoleného medoidu, kteří se budou prohledávat. Rovněž se stanoví, do jaké vzdálenosti se budou sousedé prohledávat a počet kroků celého procesu. Stejně jako v metodě CLARA se zde vychází z opakovaných pokusů. CLARANS tedy využívá mechanismu náhodného výběru uzlů a prohledávání uzlů ve stanovené vzdálenosti. Pokud při této iteraci narazí na uzel, který lépe vyhovuje, vybere ho a pokračuje z něj. V opačném případě je označen jako lokální minimum. Cyklus končí po dosažení stanovených kroků procesu. CLARANS tedy pracuje s náhodně zvolenými objekty po celou dobu procesu. Zatímco CLARA má pevně stanovený vzorek v každém kroku (Han a kol., 2006).

2.7 Validace shlukové analýzy

Pro určení výsledného počtu shluků lze využít dvou základních přístupů. Jsou jimi heuristické procedury a formální testy. Jako příklad první možnosti lze zmínit určení výsledného počtu shluků na základě dendrogramu.

Druhá možnost zahrnuje využití tzv. validačních indexů. Do těchto indexů lze řadit například izolační index, metodu siluety a Dunnův validační index. Zkoumají shlukovací metody, které jsou počítány několikrát s různými proměnnými, stanovují vhodnost metod a v neposlední řadě dovedou definovat i optimální počet shluků (Řezanková a kol., 2007).

3 Analýza vybraných logistických procesů ve zkoumaném podniku

Cílem této kapitoly je seznámení se firmou, pro kterou je bakalářská práce realizovaná. Dochází zde k nastínění vybraných logistických procesů a analýze současného stavu.

3.1 Představení společnosti

Skupina ČEZ je energetickým koncernem působícím v celkem sedmi zemích střední a jihovýchodní Evropy a v Turecku. Akciová společnost ČEZ byla založena v roce 1992 Fondem národního majetku ČR jakožto přeměna státního podniku České energetické závody. ČEZ, a. s. (dále jen ČEZ) je dnes mateřskou společností Skupiny ČEZ, která vznikla spojením distribučních společností Severočeská energetika, Severomoravská energetika, Středočeská energetika, Východočeská energetika a zmiňované akciové společnosti ČEZ. Centrála společnosti sídlí v České republice. Akcie ČEZ jsou obchodovány na pražské a varšavské burze cenných papírů. Nejvýznamnějším akcionářem mateřské společnosti ČEZ je Česká republika s podílem na základním kapitálu téměř 70 %[8]. Pro Českou republiku vykonává správu jejího akciového podílu Ministerstvo financí České republiky.

3.1.1 Předmět podnikání

Mezi primární podnikatelské činnosti se řadí výroba, nákup, distribuce a prodej elektrické energie zákazníkům všech velikostních skupin. Tyto činnosti rovněž tvoří ve společnosti ČEZ dominantní objem tržeb. V České republice je mateřská společnost ČEZ nejvýznamnějším výrobcem elektrické energie (se 75 procentním výrobním podílem) pro téměř 3,5 mil. odběrných míst.

Ze sekundárních činností lze zmínit především dodávky tepla z kombinované výroby elektřiny a tepla, zpracování vedlejších energetických produktů nebo působení na trhu telefonních služeb (ČEZ, a. s., ©2016).

3.2 ČEZ Distribuční služby, s.r.o.

ČEZ Distribuční služby, s.r.o., (dále jen ČDS), byly založeny 1. 7. 2005 jako integrovaná dceřiná společnost Skupiny ČEZ. Jako základní náplň práce ČDS lze uvést zejména komplexní

[8] Ke dni 31. 12. 2014.

zajištění služeb v oblasti provozování, odstraňování poruch, údržby a oprav distribuční soustavy.

V roce 2003 došlo ke spojení této dceřiné společnosti s dalšími významnými členy Skupiny ČEZ, konkrétně s ČEZ Měření, s.r.o. a s ČEZ Logistika, s.r.o. K činnostem ČDS tak přibyly služby související s elektroměrovou činností a skladovým hospodářstvím (ČEZ Distribuční služby, s.r.o., ©2016).

3.3 Odbor Správa měřidel

Veškerou činnost spojenou s měřicími přístroji mají na starosti dvě oddělní Správy měřidel. Jsou jimi AMS[9] a SLM[10].

Nejvýznamnější činností oddělení AMS je ověřování elektroměřících přístrojů. Ověřování probíhá jak na nových kusech přístrojů, které se dále připravují pro koncové zákazníky, tak na přístrojích, které jsou staženy ze sítě.

V případě prvně uváděných, procházejí přístroje sérií statistických pokusů, pomocí kterých probíhá ověření na náhodně vybraném vzorku. Pokud výsledný vzorek projde, je přijata i dodávka několika tisíc kusů.

Druhé zmiňované přístroje mají stanovený interval pro periodickou výměnu, jelikož jim vypršela zákonná lhůta. Elektroměry se stahují ze sítě, následně jsou posílány na příslušná pracoviště, kde je provedeno jejich třídění. Některé přístroje jsou posílány na očištění a opravu do oddělení AMS, jiné, většinou zastaralé a neopravitelné, jsou fyzicky znehodnocovány. O ekologickou likvidaci se stará externí smluvní partner.

SLM zajišťuje samotný logistický tok veškerých měřících přístrojů. Přístroje přijímá buďto nové (ty, které prošly sérií statistických testů) nebo staré (očištěné a opravené), které jsou připraveny k opětovnému použití. Přístroje dále vyskladňuje, eviduje a realizuje jejich přesun na jednotlivá pracoviště k montérům, kteří zabezpečují poslední fázi procesu. Největším pracovištěm je logistika ve Skutči, odkud se realizují rozvozy a svozy na detašované pracoviště v Liberci a v Ostravě. V Hradci Králové sídlí, kromě vedení společnosti ČDS a oddělení Správy měřidel, zaměstnanci mající za úkol evidenční nápravu technických dat přístrojů a jejich údržbu v informačním systému.

[9] Autorizované metrologické středisko a opravna.
[10] Správa a logistika měřidel.

3.4 Popis současného stavu

V současném modelu logistiky ve společnosti ČEZ probíhá v rámci ČR oboustranná komunikace mezi celkem čtyřiceti jedna městy. Jednotlivá města jsou rozdělena do následujících pěti regionů[11]:

- Sever (7) - Česká Lípa, Děčín, Liberec, Litoměřice, Výškov, Teplice, Ústí nad Labem.
- Morava (11) – Bruntál, Frýdek Místek, Havířov, Nový Jičín, Olomouc, Opava, Ostrava, Šumperk, Jeseník, Třinec, Valašské Meziříčí.
- Střed (7) – Skuteč, Benešov, Kladno, Kolín, Mělník, Mladá Boleslav, Příbram.
- Východ (8) - Česká Třebová, Havlíčkův Brod, Hradec Králové, Jičín, Náchod, Pardubice, Svitavy, Trutnov.
- Západ (8) – Domažlice, Horní Bříza, Cheb, Karlovy Vary, Klatovy, Plzeň, Stříbro, Tachov.

Hlavní, již zmiňované pracoviště se nachází ve středních Čechách, konkrétně ve městě Skuteč. Toto centrum zaměstnává největší počet pracovníků v uváděném procesu. Oproti pracovištím v Ostravě a v Liberci, kde je počet pracovníků roven čtyřem, zde nepůsobí pouze členové oddělení SLM, ale také AMS. Ze Skutče je zabezpečován svoz a rozvoz měřících přístrojů do regionů Střed, Východ a Západ. Tento logistický tok je ve firmě ČEZ uskutečňován vlastní dopravou. Firma má pro tyto účely k dispozici jedno nákladní vozidlo značky Mercedes-Benz s maximální nosností 5275 kg. K rozvozové činnosti jsou přiřazeni dva řidiči, kteří se dle potřeby střídají. V některých případech, konkrétně pro rozvoz a svoz mezi městy v Západočeském regionu, je nutno mít k dispozici oba dva zmiňované zaměstnance. Druhým omezením, se kterým je nutno počítat je celková velikost svozů a rozvozů, které se napříč služebnami liší. Veškeré přístroje se sváží a rozváží v uzamčených v přepravnících neboli kontejnerech, které jsou k této činnosti přizpůsobené. Elektroměry nemají stejné rozměry a váhy a z tohoto důvodu se jich do kontejneru vleze různé množství napříč vybranými druhy. Pro vozidlo z města Skuteč je maximální počet kontejnerů stanoven na 17 kusů, pro města Ostravu a Liberec na 16 kusů.

V popisu současného stavu se zaměřuji na rozvozové trasy, které jsou platné v roce 2016. Je důležité podotknout, že v průběhu minulých let se jednotlivé trasy obměňovaly podle

[11] Rozdělení vyplívá z firemních údajů, nekopíruje přesně geografickou polohu měst.

v té době aktuálního plánu pro sběr a výměnu elektroměrů v daných lokalitách. Co se týče porovnání s rokem 2015, došlo v lednu a v dubnu toho roku ke zrušení závozních míst Brandýs nad Labem a Vsetín.

Dopravu ze Skutče lze rozdělit na dvě části. V první z nich se nachází detašované pracoviště v Ostravě a Liberci. V rámci této části jsou zmíněná města zásobována přístroji, které se zde připravují pro jednotlivá města v regionech. Primárně tedy nezabírají kontejnery hlavní část nákladu. Ta je tvořena koši, ve kterých jsou elektroměry uskladněny. Následně probíhá vychystávaní měřících přístrojů stejně jako ve Skutči. V rámci každé trasy je uvedena i její průměrná kilometrová vzdálenost vyplívající z firemních informací.

a) Zásobování detašovaných pracovišť (probíhá 2x za měsíc):
   - Skuteč - Ostrava – Skuteč (438 km)
   - Skuteč – Liberec – Skuteč (316 km)

Druhá část dopravy ze Skutče zahrnuje již zmiňovaný rozvoz do jednotlivých měst, které jsou tímto městem obsluhovány.

b) Rozvoz a svoz (každá trasa se jezdí jednou měsíčně):
   - Skuteč – Horní Bříza – Plzeň - Stříbro – Skuteč (631 km)
   - Skuteč – Benešov – Havlíčkův Brod – Skuteč (291 km)
   - Skuteč – Kolín – Skuteč (190 km)
   - Skuteč – Klatovy - Domažlice – Tachov – Skuteč (709 km)
   - Skuteč – Mladá Boleslav - Mělník – Skuteč (369 km)
   - Skuteč – Hradec Králové – Pardubice – Skuteč (139 km)
   - Skuteč – Náchod – Trutnov – Skuteč (241 km)
   - Skuteč – Kladno – Skuteč (394 km)
   - Skuteč – Příbram – Skuteč (420 km)
   - Skuteč – Karlovy Vary - Cheb – Skuteč (702 km)
   - Skuteč – Svitavy - Česká Třebová – Skuteč (121 km)

V případě města Olomouce probíhá rozvoz a svoz v jednom (nejčastěji prvním) termínu při zásobování Ostravy. Rozvoz a svoz města Jičín probíhá rovněž jednou (nejčastěji v druhém) termínu při zásobování Liberce.

Rozvozy a svozy do jednotlivých měst ve Východočeském a Moravskoslezském regionu zajišťují externí firmy. V tomto případě je dopravcům účtováno za nakládku a vykládku dodávaného zboží a za maximální počet ujetých kilometrů v rámci každé trasy. Maximální počet kilometrů je nastaven z důvodu neočekávaných dopravních situací. Každá trasa se rovněž uskutečňuje jednou za měsíc a pro severní Moravu jsou celkově tři:

- Ostrava – Opava – Bruntál – Jeseník – Šumperk – Ostrava (387 km)
- Ostrava – Nový Jičín – Valašské Meziříčí – Ostrava (178 km)
- Ostrava – Havířov - Třinec - Frýdek Místek – Ostrava (148 km)

V případě východních Čech jsou stejně jako pro severní Moravu trasy tři:

- Liberec – Ústí nad Labem - Teplice – Liberec (240 km)
- Liberec – Česká Lípa – Děčín – Liberec (208 km)
- Liberec - Litoměřice – Výškov – Liberec (334 km)

4 Návrh a implementace optimalizačního modelu v podniku

V této části práce bude pomocí metod shlukové analýzy nastíněno nové rozdělení měst. V práci bude využit program IBM SPSS Statistics.

4.1 Představení softwaru IBM SPSS Statistics

IBM SPSS Statistics je softwarová platforma využívaná k podpoře rozhodování jednotlivců, systémů i podniků. Patří mezi celosvětově rozšířené statistické systémy pro aplikace ve vědě, marketingu, laboratorních měřeních a rovněž pro sumarizaci dat z velkých i menších databází různého typu. Ve všech zmiňovaných odvětvích poskytuje celou řadu pokročilých technik a algoritmů. Skládá se z velkého počtu modulů, z nichž každý má v programu svůj speciální účel. Základním modulem je IBM SPSS Statistics Base, který umožňuje export a import dat mnoha formátů s využitím ODBC[12], datové manipulace, transformace dat, statistické přehledy aj. Popis ostatních modulů je dostupný v Petr (2012).

Vstupní datový soubor je podrobně popsán z hlediska kódů kategorií a proměnných, přičemž základní prostředí je postupně tvořeno

- datovým editorem, představující tabulkový editor pro prohlížení a úpravu dat a
- výsledným oknem, které je rozděleno na části pořadač výstupů a textový editor výstupů. Na druhou zmiňovanou část navazují speciální okna pro úpravu výsledných grafů a tabulek.

Pro shlukovací analýzu nabízí SPSS práci s metodami

- hierarchického shlukování (uvedené v kapitole 2.4.1),
- *k*-průměrů (uvedenou v kapitole 2.5.1),
- dvoukrokové shlukovací analýzy (uvedené v kapitole 2.6.2),

které mohou zpracovávat kvantitativní i kvalitativní data.

Licence softwaru IBM SPSS Statistics ve verzi 23 je volně dostupná pro zaměstnance a studenty VŠB-TU. Dalším často využívaným programem pro shlukovou analýzu je software STATISTICA, který obsahuje v oblasti shlukovacích metod procedury pro hierarchické shlukování a proceduru pro shlukování pomocí *k*-průměrů.

[12] Standartní aplikační rozhraní pro přístup k datům.

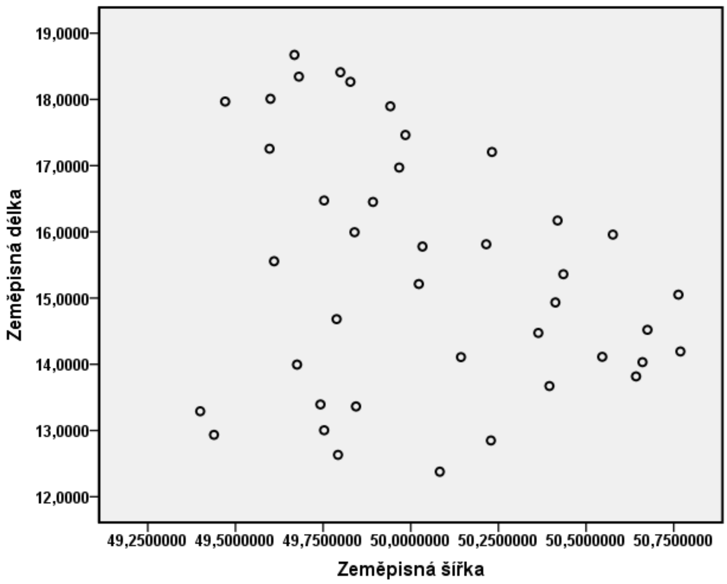
4.2 Stanovení parametrů optimalizace logistických procesů

Přípravná fáze je základem každého shlukovacího procesu. Analytik musí datům porozumět, popřípadě ošetřit některé nežádoucí vlivy.

4.2.1 Analýza vstupních dat

V této práci se vychází ze vstupního datového souboru zobrazeného v Příloze č. 1., který obsahuje informace o geografické poloze jednotlivých pracovišť ČDS. Datová matice je tvořena názvy měst a proměnnými obsahující zeměpisnou šířku a zeměpisnou délku. Samotná data byla získána z internetového serveru mapy.cz. Pro posuzování shluků byly vybrány záměrně pouze tyto dvě proměnné. Ostatní zvažované proměnné (například průměrný počet kontejnerů k přepravě), budou využity v další části práce. V této části by pouze zkreslily výsledek jednotlivých shluků, protože by mohly shlukovat k sobě objekty podle kritérií, která nejsou v modelovém příkladu podstatná.

Prvním grafickým výstupem je prezentace vstupních dat (obrázek 4.1), tedy jejich rozložení ve dvourozměrném prostoru. Zejména u dat popisující polohu jakýchkoli objektů je tento prvotní grafický přehled důležitou součástí analýzy.



Obrázek 4.1 - Grafické rozložení analyzovaných dat, zdroj: Vlastní zpracování.

Zjištění chybějících a odlehlých hodnot

Analýza úplnosti datového souboru tvoří základ celé přípravy. Analytik si musí být vědom, zda soubor obsahuje všechna data potřebná ke shlukování. V případě většího počtu dat (řádově v tisících) je vhodné tuto analýzu provést pomocí uzlu „Missing Value Analysis“

(tabulka 4.1). Počet chybějících hodnot a jejich procentuální zastoupení v datovém souboru zobrazuje sloupec „Missing“. Dále tato analýza dokáže podle kritéria stanoveného pod tabulkou odhalit odlehlá pozorování (sloupec „No. of Extremes“), která můžou do značné míry zkreslovat variabilitu[13] dat jednotlivých proměnných. Pokud je v datovém souboru variabilita vysoká, značí to velkou vzájemnou odlišnost hodnot dané proměnné. Zároveň vysoká variabilita signalizuje, že vybrané statistické charakteristiky souboru (modus, medián, střední hodnota) nejsou vhodnými charakteristikami hodnot dané proměnné. Výsledek shlukování proto mohou odlehlé hodnoty silně ovlivňovat.

Druhou možností, jak odlehlá pozorování odhalit, je provedení hierarchických shlukovacích algoritmů. Pokud jsou tyto objekty v datech zastoupeny ve větším množství, budou tvořit jednoobjektové shluky, zatímco ostatní objekty budou mít tendenci se seskupovat do shluků o více objektech.

Jak v případě chybějících, tak odlehlých hodnot je nutné tato data vynechat z datového souboru.

Tabulka 4.1 - Zjištění chybějících a odlehlých hodnot ze vstupní datové matice

	N	Mean	Std. Deviation	Missing		No. of Extremes[a]	
				Count	Percent	Low	High
Z_S	41	50,041041	0,393906	0	,0	0	0
Z_D	41	15,382018	1,856827	0	,0	0	0

a. Number of cases outside the range (Q1 - 1.5*IQR, Q3 + 1.5*IQR)

Dalším možným řešením pro stanovení odlehlých pozorování je grafická analýza dat vstupních proměnných (viz obrázek 4.1).

Popisná statistika

Popisnou statistiku lze zobrazit pomocí uzlu „Frequencies“. Pro jednotlivé proměnné jsou vypočteny statistické charakteristiky jako například průměr, standardní odchylka od průměru, medián, rozptyl, variační rozpětí, směrodatná odchylka, minimum, maximum (tabulka 4.2). Například hodnota variačního rozpětí („Range“) je vyšší pro zeměpisnou délku, což z geografického hlediska odpovídá České republice. Rozdíl mezi sledovanými hodnotami s největší a nejmenší zeměpisnou délkou je vyšší než v případě rozdílu maximální a minimální zeměpisné šířky. Z deskriptivních charakteristik se vychází v další části fáze analýzy dat.

[13] Variabilita značí proměnlivost souboru – jak moc si jsou nebo nejsou data podobná.

Tabulka 4.2 - Základní deskriptivní statistiky

	Zeměpisná šířka	Zeměpisná délka
N Valid	41	41
Mean	50,041041339	15,382017676
Std. Error of Mean	,0615177041	,2899877264
Median	49,966983100	15,211983900
Std. Deviation	,3939055022	1,8568274398
Variance	,155	3,448
Range	1,3698806	6,2939333
Minimum	49,3993294	12,3779092
Maximum	50,7692100	18,6718425

Korelace

Při posuzování závislosti mezi jednotlivými proměnnými obsahující pouze kvantitativní data lze vycházet z Pearsonova korelačního koeficientu (2.10). V modelovém případě se jeho hodnota rovná $r_{kl} = -0{,}248$, což značí slabou negativní korelaci mezi posuzovanými proměnnými. Na základě hodnoty korelačního koeficientu lze proměnné považovat za nezávislé.

Standardizace

V přípravné fázi je rovněž žádoucí řešit otázku, zda je, či není potřeba vstupní data standardizovat. Ve vstupní datové matici se nacházejí proměnné obsahující hodnoty o stejných jednotkách. Z tohoto důvodu není standardizace v případě řešeného problému nutná. Dále však platí, že proměnné s vyšší mírou proměnlivosti (směrodatnou odchylkou) mají větší vliv na míru podobnosti (kapitola 2.1.2). Z tabulky 4.2 je patrná různá variabilita v rámci datové matice. Z tohoto důvodu byly vstupní proměnné standardizovány na stejnou škálu s jednotkovým rozptylem a nulovým průměrem. Standardizované hodnoty vstupních proměnných jsou ve vstupní datové matici zobrazeny ve sloupcích ZZ_S a ZZ_D. Hodnoty z těchto dvou sloupců jsou použity v metodě *k*-průměrů. Pro hierarchickou shlukovou analýzu je v SPSS možné tuto úpravu provést přímo v rámci hierarchických shlukovacích metod.

Stanovení počtu shluků

Podle povahy analyzovaných dat lze v našem případě odhadnout přinejmenším rozsah, ve kterém se počet výsledných shluků bude pohybovat. Z celkového počtu objektů lze usuzovat, že pro firmu bude nejvíce reálné uvažovat o celkovém počtu shluků v rozmezí od 2 po maximálně 5 nebo 6 shluků. V případě vyššího počtu shluků by vzdálenosti mezi centry byly natolik malé, že rozvoz a svoz objektů v jednom shluku by se dal v přijatelném časovém intervalu zvládnout z centra shluku jiného.

Výsledné určení shluků je v práci založeno na heuristice (kapitola 2.7). Ke konečnému rozložení objektů do shluků byla využita zejména analýza aglomerativní tabulky a dendrogramů.

4.3 Porovnávání a výpočet vybraných metod v prostředí SPSS

V první fázi shlukování jsou v práci využity hierarchické metody, popsané v kapitole 2.4.1. Jmenovitě se jedná o metodu

- nejbližšího souseda,
- nejvzdálenějšího souseda,
- průměrné vazby pro vnitroshlukovou vzdálenost,
- centroidní,
- mediánovu,
- Wardovu.

V případě prvních třech metod se pro stanovení vzdálenosti vychází z euklidovské matice vzdáleností (tabulka 4.3). U zbylých dvou, Wardovy shlukovací metody a centroidní metody, je míra vzdálenosti stanovena pomocí čtvercové euklidovské vzdálenosti (tabulka 4.4). Hodnoty představují míru podobnosti mezi všemi objekty z datové matice. Čím je hodnota mezi dvěma objekty menší, tím jsou si objekty podobnější (bližší). Matice vzdáleností je každým krokem přepočítávána. To, jakým způsobem jsou spojeny objekty do dvou a následně více shluků, určuje vybraná shlukovací metoda. Kupříkladu pokud je zvolena metoda nejbližších sousedů, dochází v matici k nalezení dvou bodů, které jsou si neblíže (hodnota vzdálenosti je nejmenší) a dochází k jejich spojení a stanovení nové matice, respektive k úpravě stávající o nově vzniklé shluky. Postup je neustále opakován, dokud nejsou všechny objekty přiřazeny do shluků.

Tabulka 4.3 - Ukázka matice euklidovských vzdáleností

Case	Euclidean Distance					
	1:Česká Lípa	2:Děčín	3:Liberec	4:Litoměřice	5:Výškov	6:Teplice
1:Česká Lípa	,000	,297	,363	,395	,844	,388
2:Děčín	,297	,000	,462	,568	,990	,380
3:Liberec	,363	,462	,000	,748	1,193	,732
4:Litoměřice	,395	,568	,748	,000	,450	,292
5:Výškov	,844	,990	1,193	,450	,000	,632
6:Teplice	,388	,380	,732	,292	,632	,000

Tabulka 4.4 - Ukázka matice čtvercové euklidovské vzdálenosti

Case	Squared Euclidean Distance					
	1:Česká Lípa	2:Děčín	3:Liberec	4:Litoměřice	5:Výškov	6:Teplice
1:Česká Lípa	,000	,088	,132	,156	,713	,150
2:Děčín	,088	,000	,213	,323	,979	,145
3:Liberec	,132	,213	,000	,560	1,424	,535
4:Litoměřice	,156	,323	,560	,000	,202	,085
5:Výškov	,713	,979	1,424	,202	,000	,400
6:Teplice	,150	,145	,535	,085	,400	,000

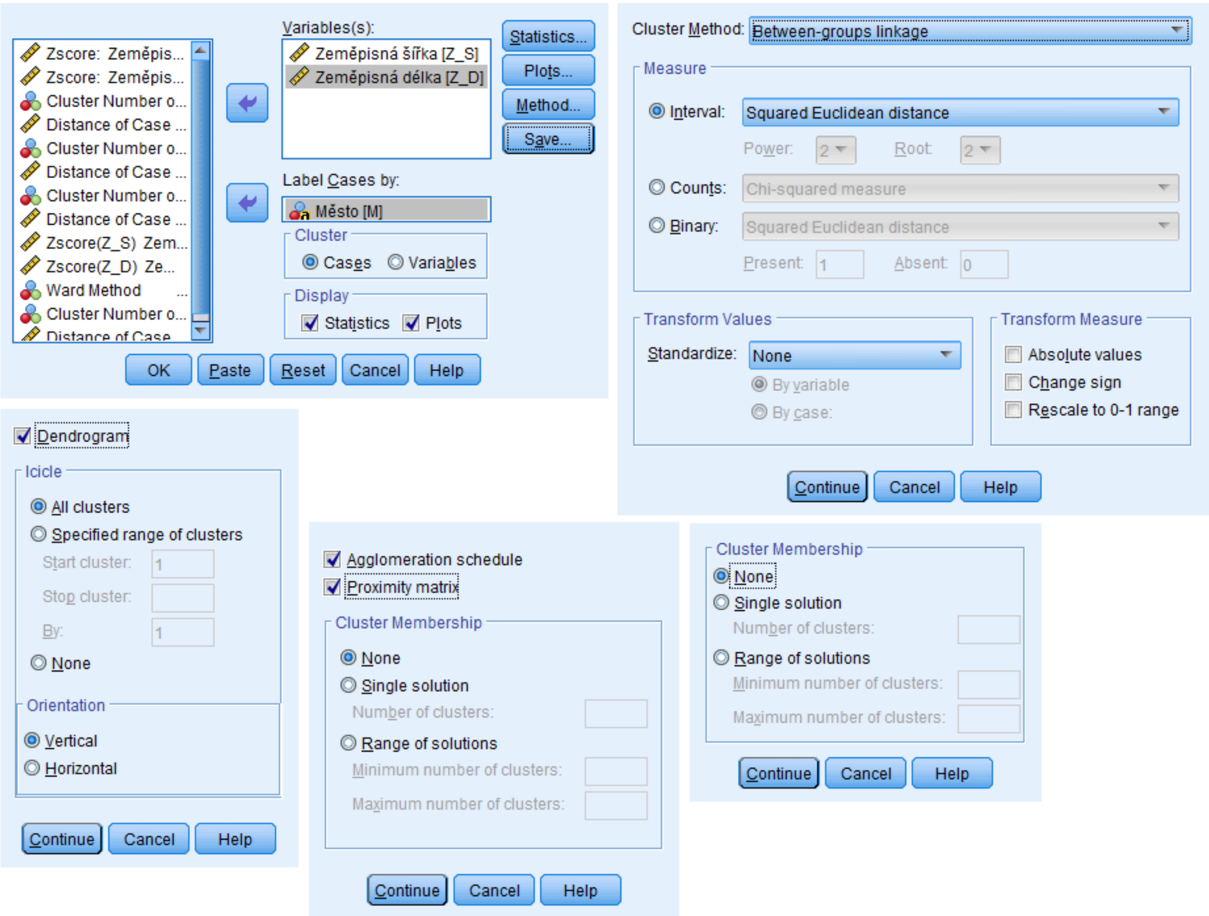
4.3.1 Shlukování pomocí hierarchických metod

Ve všech hierarchických metodách lze proměnné standardizovat podle určitého kritéria (například podle variačního rozpětí 0 - 1). Zde je využita možnost standardizace pomocí *z*-skóre, která je stejná pro každou hierarchickou metodu. Ukázka SW prostředí SPSS při využití hierarchického shlukování je zobrazena na obrázku 4.2.

Základní nastavení shlukování zahrnuje pro každou metodu volitelné zobrazení statistických možností (tzv. „Agglomeration Schedule“, „Proximity Matrix“) a grafické vizualizace (dendrogram, krápníkový graf). Dále je zde nastaveno kritérium shlukování. K volbě je možnost shlukovat podle proměnných nebo podle objektů. V práci je využito shlukování podle objektů, jelikož jsou to právě ony, které chceme shlukovat. K hodnotám analyzovaných proměnných jsou přiřazeny odpovídající města z prvního sloupce. V uzlu „Method“ (při výběru shlukovací metody a metriky vzdálenosti) je možnost provést, kromě výše uvedené standardizace, transformaci dat na absolutní hodnoty nebo stanovit jejich rozsah od nuly po jedničku.

Nicméně z důvodu rozsáhlosti metod a také jejich částečné shodě, jsou v práci popsány pouze některé. K jejich detailnější prezentaci slouží přiložené samostatné sešity MS Excel. Jednotlivé listy souboru obsahují

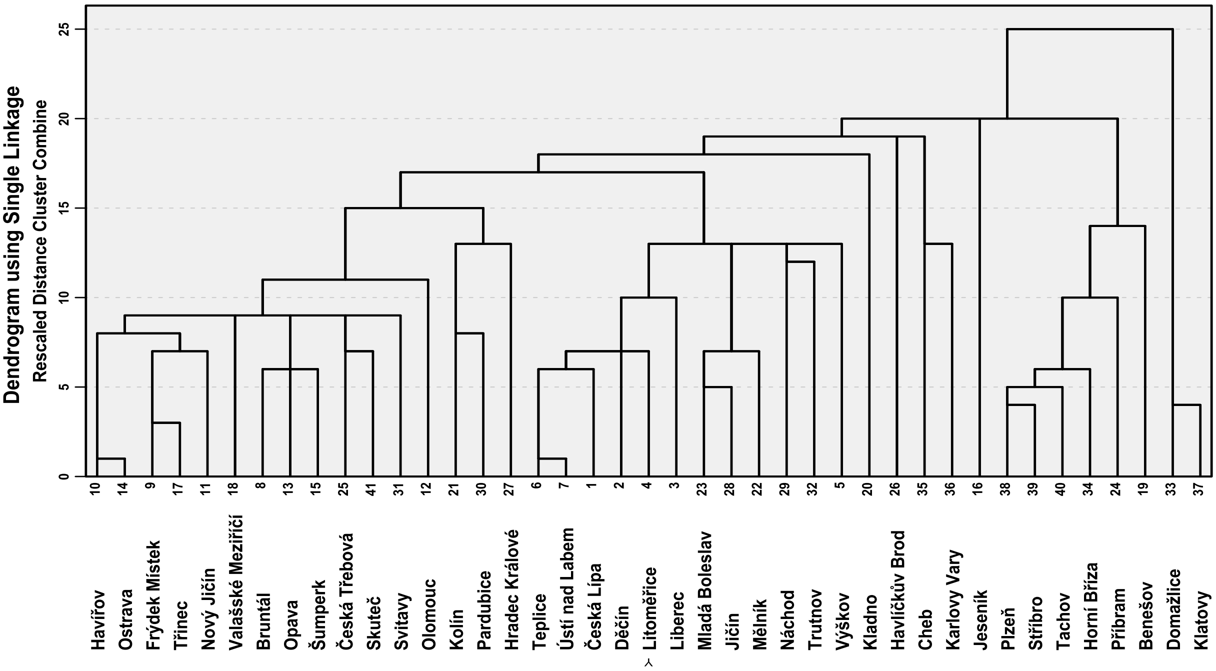
- vstupní data,
- matici vzdáleností,
- postup a výsledné rozložení shlukovacího procesu,
- dendrogram.



Obrázek 4.2 - Prostředí hierarchických metod v SPSS, zdroj: Vlastní zpracování.

4.3.2 Metoda nejbližšího souseda

Hned první metoda, která byla v praktické části použita, dosahovala obtížně interpretovatelných výsledků. Jak je z výsledného dendrogramu (obrázek 4.3) patrné, dochází mezi shlukovanými objekty k řetězení. Výsledné shluky jsou proto velice špatně odhalitelné. Objekty se neustále přidávají do jednoho shluku, který v souboru dat dominuje. Pokud tedy rozdělíme datový soubor do 6 shluků, mohou se vyskytnout i shluky jednoobjektové. Analytik by posléze musel rozhodnout, zda tyto objekty zahrnovat do odlehlých shluků popřípadě je eliminovat ze vstupních dat. Nicméně v dalších metodách již tento problém nenastával.



Obrázek 4.3 - Dendrogram metody nebližšího souseda, zdroj: Vlastní zpracování.

4.3.3 Metoda nejvzdálenějšího souseda

Pro další porozumění problematice jednotlivých kroků shlukování je nutné v první řadě popsat výslednou aglomerativní tabulku („Agglomeration Schedule“), jejíž poslední kroky jsou znázorněny pomocí tabulky (4.6).

1. První sloupec („Stage“) značí počet kroků shlukovacího procesu, který je obvykle roven *n*-1. Na začátku se vychází z aglomerativního přístupu (kapitola 2.4.1), kdy všechny objekty tvoří samostatný shluk. V posledním kroku jsou všechny analyzované objekty spojeny do jednoho shluku.
2. Sloupce „Cluster Combined“ ukazují, které shluky byly v daném kroku zkombinovány.
3. Prostřední sloupec „Coefficients“ znázorňuje koeficienty podobnosti neboli hodnoty vzdáleností mezi dvěma shluky z druhého kroku. Jejich hodnoty jsou dány metodou, která byla definována při vytváření shluků (pokud je například zvolena metoda nejkratší vzdálenosti, bude tento sloupec obsahovat hodnotu mezi dvěma nejbližšími objekty ve shluku). Koeficienty podobnosti obsažené v tomto sloupci rovněž slouží jako přibližný přehled počtu shluků. Sleduje se zejména velký rozdíl mezi

hodnotami koeficientů dvou po sobě jdoucích kroků. Příliš velký skok udává různorodost mezi dvěma shluky a naznačuje přerušení shlukovacího algoritmu. Zajímá nás tedy krok, ve kterém došlo k výraznému skoku dvou následujících hodnot koeficientů podobnosti.

4. „Stage Cluster First Appears“ značí kolo, ve kterém se shluk poprvé objevil.
5. Sloupec „Next Stage“ zobrazuje kolo, ve kterém daný shluk objeví znova a zkombinuje se s jiným shlukem.

Pro popis procesu shlukování s využitím metody nejvzdálenějšího souseda jsou použita identifikační čísla jednotlivých měst, uvedená v Příloze č. 1. Tabulka 4.5 obsahuje všechny kroky shlukovacího procesu až do spojení objektů v jeden jediný shluk. Z této tabulky je možné vysledovat poslední objekt, který se spojuje do shluků obsahující dva a více objektů. Tímto objektem je Havlíčkův Brod a ke sloučení dochází až ve 29. kroku. Toto město se tím pádem stává nejvzdálenějším samostatným shlukem při využití shlukování pomocí metody nejvzdálenějšího souseda.

Tabulka 4.5- Jednotlivé kroky metody nejvzdálenějšího souseda

Číslo kroku	Shluk	Číslo kroku	Shluk	Číslo kroku	Shluk
1.	10, 14	15.	29, 32	29.	19, 24, 26
2.	6, 7	16.	38, 34, 40, 39	30.	9, 17, 10, 14, 11, 18, 12
3.	9, 17	17.	5, 22	31.	33, 37, 38, 34, 40, 39
4.	38, 39	18.	35, 36	32.	21, 30, 27, 25, 41, 31
5.	33, 37	19.	9, 17, 10, 14	33.	5, 22, 20, 35, 36
6.	23, 28	20.	1, 2, 3	34.	1, 2, 3, 6, 7, 4, 23, 28, 29, 32
7.	8, 13	21.	19, 24	35.	8, 13, 15, 16, 21, 30, 27, 25, 41, 31
8.	25, 41	22.	11, 18, 12	36.	19, 24, 26, 33, 37, 38, 34, 40, 39
9.	6, 7, 4	23.	8, 13, 15	37.	8, 13, 15, 16, 21, 30, 27, 25, 41, 31, 9, 17, 10, 14, 11, 18, 12
10.	1, 2	24.	21, 30, 27	38.	1, 2, 3, 6, 7, 4, 23, 28, 29, 32, 5, 22, 20, 35, 36
11.	38, 39, 34	25.	5, 22, 20	39.	8, 13, 15, 16, 21, 30, 27, 25, 41, 31, 9, 17, 10, 14, 11, 18, 12, 19, 24, 26, 33, 37, 38, 34, 40, 39
12.	21, 30	26.	23, 28, 29, 32	40.	Všechny objekty tvoří 1 shluk
13.	11, 18	27.	1, 2, 3, 6, 7, 4		
14.	25, 41, 31	28.	8, 13, 15, 16		

Stanovení počtu shluků

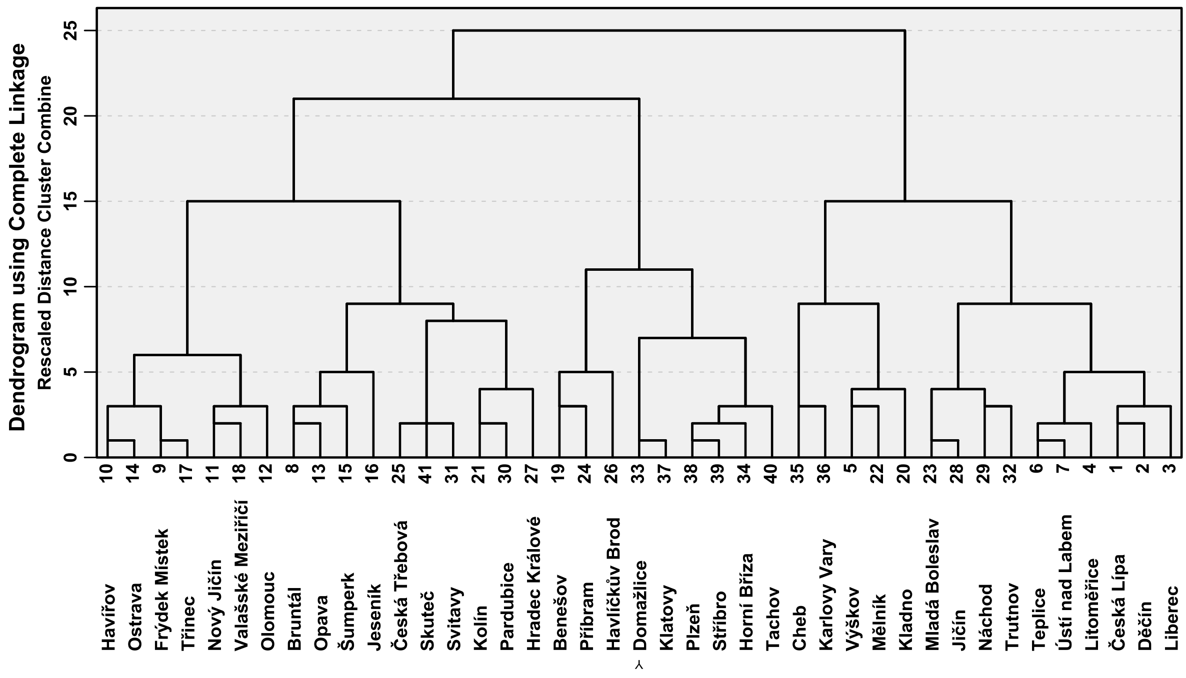
Pro úplnost jsou znázorněny poslední kroky aglomerativní tabulky (tabulka 4.6). Podle výše popsaného postupu lze z výsledné tabulky určit celkový počet shluků. První výrazné zvýšení hodnot koeficientů lze sledovat mezi 36. a 37. krokem shlukovacího procesu. Výsledný počet shluků se rovná rozdílu mezi počátečním počtem objektů a krokem, ve kterém dochází k výraznému skoku v hodnotách koeficientů. V některých případech může nastat problém tento výrazný skok odhalit. Ke grafické prezentaci je využit bodový graf (k nahlédnutí pouze v excelovských souborech), který na ose *x* zobrazuje jednotlivé kroky algoritmu a na ose *y* jim příslušné hodnoty koeficientů. Jsou zde lépe patrné návaznosti dvou po sobě jdoucích kroků. V případě metody nejvzdálenějšího souseda je výsledná hodnota počtu shluků k rovna 5.

Tabulka 4.6 - Aglomerativní tabulka metody nejvzdálenějšího souseda

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
34	1	23	1,389	27	26	38
35	8	21	1,460	28	32	37
36	19	33	1,641	29	31	39
37	8	9	2,218	35	30	39
38	1	5	2,300	34	33	40
39	8	19	3,268	37	36	40
40	1	8	3,873	38	39	0

Po opětovném provedení shlukovací metody nejvzdálenějšího souseda a využití volby pro rozdělení objektů do výsledných pěti shluků, je vytvořen nový sloupec s proměnnou obsahující čísla shluků, ke kterým jsou jednotlivé objekty přiřazeny. Tato možnost je vhodná v případě, kdy známe výsledný počet shluků a potřebujeme už je pouze efektivněji přiřadit k objektům.

Jak již bylo zmíněno, výstupy uváděné shlukovací metody se rovněž dají odvodit z dendrogramu (obrázek 4.4) pomocí analýzy největších skoků. Horizontální tečkované čáry dendrogramu indikují vzdálenost v upraveném měřítku (v rozmezí od 0 do 25), do kterého jsou shluky formovány. Tato vzdálenost se nazývá hladina spojování. Pokud by byl proveden horizontální řez na hladině spojování 10, bylo by patrné rozdělení objektů do 6 shluků. Nejpodobnější jsou si ty objekty, které jsou sloučeny na nejnižší hladině spojování. Z hlediska celkového výsledku je z grafu patrné rozdělení na pět výše uvedených shluků. Rovněž je zde přijatelná varianta pro stanovení tří shluků na hladině spojování 20.



Obrázek 4.4 - Dendrogram metody nejvzdálenějšího souseda, zdroj: Vlastní zpracování.

Na základě zmíněných skutečností jsou města rozdělena do pěti stanovených shluků podle tabulky 4.7. Z konečného rozdělení lze pozorovat, že počet objektů je proměnlivý. Zejména u druhého shluku je patrné mále zastoupení objektů v porovnání s ostatními shluky.

Tabulka 4.7- Výsledné zobrazení shlukovacího procesu metody nejvzdálenějšího souseda (v závorce je uveden počet objektů ve shlucích)

1. shluk (10)	2. shluk (5)	3. shluk (10)	4. shluk (7)	5. shluk (9)
Česká Lípa	Výškov	Bruntál	Frýdek Místek	Benešov
Děčín	Kladno	Opava	Havířov	Příbram
Liberec	Karlovy Vary	Šumperk	Nový Jičín	Havlíčkův Brod
Litoměřice	Mělník	Jeseník	Valašské meziříčí	Domažlice
Teplice	Cheb	Kolín	Olomouc	Horní bříza
Ústí nad Labem		Pardubice	Třinec	Klatovy
Mladá Boleslav		Hradec Králové	Ostrava	Plzeň
Jičín		Skuteč		Stříbro
Náchod		Svitavy		Tachov
Trutnov		Česká Třebová		

4.3.4 Wardova metoda

Wardova metoda patří k nejpoužívanější metodě hierarchického shlukování. Dosahuje totiž v porovnání s ostatními metodami nejlepších výsledků (Meloun a Militký, 2004). V matici vzdáleností je využita čtvercová euklidovská metrika (kapitola 2.1.1).

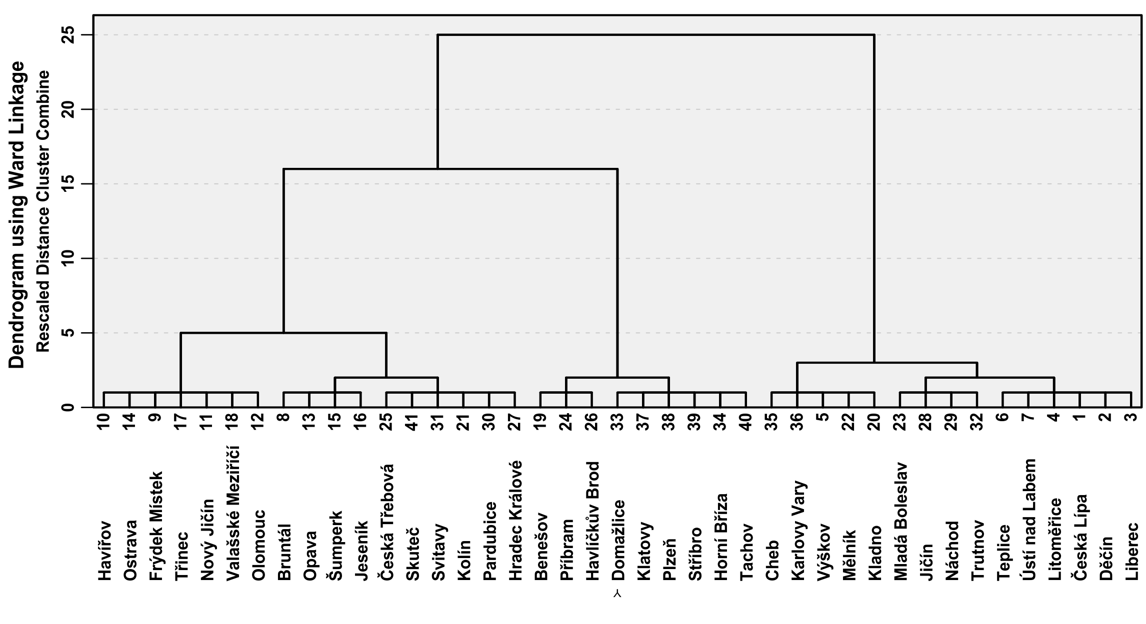
Stanovení počtu shluků

Wardova metoda je založena na rozdílném přístupu, než výše popsané metody. Z tohoto důvodu se dá očekávat jednak rozdílnost ve shlukovacím postupu, tak i v hodnotách a skocích mezi koeficienty, které jsou využity ke stanovení výsledného počtu shluků. V porovnání s metodou nejvzdálenějšího souseda se změny ve spojení shluků projevují od 9. kroku. Nicméně zásadní větší skoky v hodnotách koeficientů je možno pozorovat mezi 38. a 39. krokem shlukovacího procesu (tabulka 4.8).

Tabulka 4.8 - Aglomerativní tabulka Wardovy metody

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
34	8	21	8,168	29	31	38
35	19	33	9,742	28	33	39
36	1	23	11,669	27	26	37
37	1	5	15,941	36	32	40
38	8	9	21,819	34	30	39
39	8	19	44,436	38	35	40
40	1	8	80,000	37	39	0

Výsledný dendrogram (obrázek 4.5) je v porovnání s ostatními metodami výrazně rozdílný. V horizontálním řezu tohoto stromového grafu lze vidět, že hned na počátku bylo vytvořeno 8 shluků, které obsahovaly všechny shlukované objekty. Následně jsou z grafu patrné kroky, ve kterých docházelo k postupnému spojení objektů do pěti až jednoho shluku.



Obrázek 4.5 - Dendrogram Wardovy metody, zdroj: Vlastní zpracování.

V případě opětovného provedení Wardovy metody s využitím rozčlenění objektů do zvolených tří shluků lze pozorovat, že dva shluky jsou tvořeny téměř totožným počtem měst a ve třetím jsou obsažena města ze Západočeského kraje (tabulka 4.9). V tomhle případě zde neplatí poznatek z kapitoly 2.4.1, že výsledné shluky jsou tvořeny relativně stejným počtem objektů.

Tabulka 4.9 - Výsledné zobrazení shlukovacího procesu Wardovy metody

1. shluk (15)	2. shluk (17)	3. shluk (9)
Česká Lípa	Bruntál	Benešov
Děčín	Frýdek Místek	Příbram
Liberec	Havířov	Havlíčkův Brod
Litoměřice	Nový Jičín	Domažlice
Výškov	Olomouc	Horní Bříza
Teplice	Opava	Klatovy
Ústí nad Labem	Ostrava	Plzeň
Kladno	Šumperk	Stříbro
Mělník	Jeseník	Tachov
Mladá Boleslav	Třinec	
Jičín	Valašské Meziříčí	
Náchod	Kolín	
Trutnov	Česká Třebová	
Cheb	Hradec Králové	
Karlovy Vary	Pardubice	
	Svitavy	
	Skuteč	

Analýzu hierarchických metod zbývá doplnit ještě o poslední neuváděný výstup. Tím je jednofaktorová analýza rozptylu tzv. „ANOVA“ (tabulka 4.10). Základem je stanovení testovacího *F*-kritéria, které značí poměr meziskupinového rozptylu k rozptylu vnitroskupinovému. Na základě výsledků této metody lze zamítnout hypotézu o shodě rozptylů a tím i nulovou hypotézu analýzy rozptylu, že střední hodnoty sledovaných skupin se neliší.

Tabulka 4.10 - Jednofaktorová analýza rozptylu Wardovy metody

		Sum of Squares	df	Mean Square	F	Sig.
Z_S	Between Groups	4,636	2	2,318	56,098	,000
	Within Groups	1,570	38	,041		
	Total	6,206	40			
Z_D	Between Groups	97,576	2	48,788	45,962	,000
	Within Groups	40,336	38	1,061		
	Total	137,912	40			

4.3.5 Ostatní hierarchické shlukovací metody

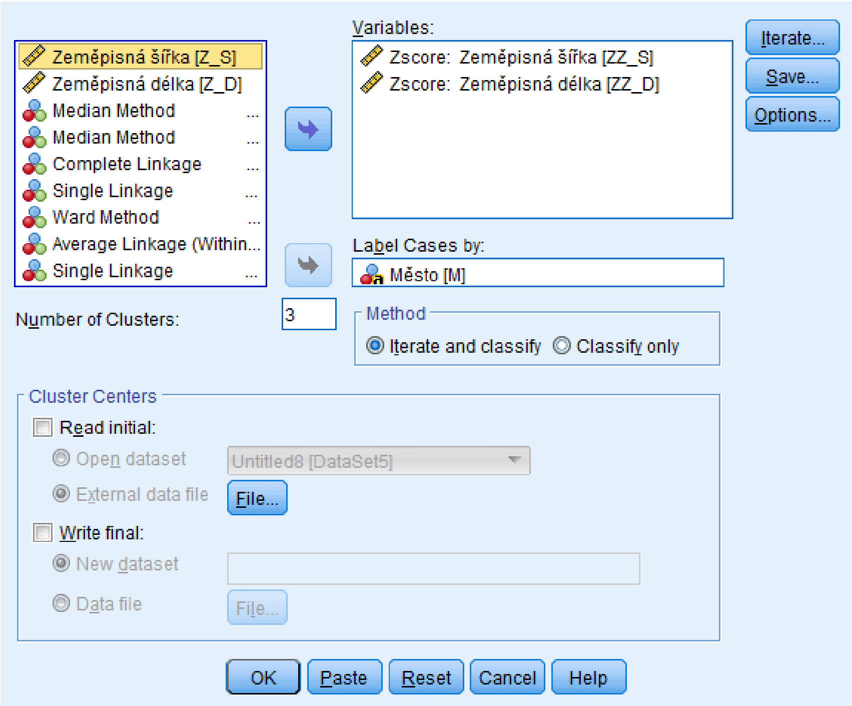
Výsledky Wardovy metody a metody nejbližšího souseda jsou zobrazeny v samostatných sešitech MS Excel. Dále je v této příloze zobrazena metoda průměrné vazby pro vnitroshlukovou vzdálenost. V případě této metody je velice dobře patrné rozdělení na tři shluky. V porovnání s výsledky Wardovy metody došlo k přeřazení Karlových Varů a Chebu do třetího shluku a Havlíčkova Brodu do shluku druhého.

Ostatní metody uvedené na začátku kapitoly 4.3 byly sice analyzovány, ale jejich výstupy (kromě dendrogramu metody nejbližšího souseda) nejsou součástí práce ani příloh. Nejméně jednoznačných výsledků dosahovala mediánová metoda. Výsledné shluky byly hůře odhalitelné, jelikož skoky v koeficientech podobnosti byli zanedbatelné. A při opětovné analýze s využitím čtvercové euklidovské metriky, metoda přiřadila na jednu stranu Šumperk ke Slezskému regionu, na druhou stranu ale Jeseník spojila s městy z východních Čech.

4.3.6 Metoda *k*-průměrů

Základní shlukovací principy metody *k*-průměrů byly popsány v kapitole 2.5.1. Oproti výše uvedeným metodám je základním předpokladem realizace této metody stanovení předem konečného počtu shluků. Pro tento výběr byly brány v potaz výsledky předešlých hierarchických metod. Tento způsob stanovení počátečních shluků je rovněž doporučován. V metodě bylo postupně voleno počáteční rozdělení na tři, čtyři nebo pět shluků. Pokud zavrhneme možnost dělení objektů do jednoho a dvou shluků, byly patrné právě u těchto tří rozložení největší skoky v rámci aglomerativní tabulky a výsledných dendrogramů.

Počáteční nastavení vlastností metody *k*-průměrů v SPSS zobrazuje obrázek 4.6. V první fázi je zvoleno, které proměnné budou shlukovány a podle čeho budou tříděny. V této metodě byly za proměnné dosazeny sloupce ZZ_S a ZZ_D. V poli „Number of Clusters“ je zvolen předem stanovený počet shluků. Důležité je zmínit, že i když jsou objekty na začátku procesu přiřazeny k určitému shluku, můžou být v konečné fázi ve shluku jiném. V uzlu „Iterate“ lze určit maximální počet iteračních kroků (přiřazování objektů k centroidům), jehož výchozí hodnota je deset. Dále je možnost iterační proces ukončit kritériem konvergence, které specifikuje míru příslušnosti objektů k centrům. Proces je ukončen tehdy, pokud je zvolená míra v dané iteraci menší než zvolená hodnota. Velmi důležitou volbou je zvolení možnosti „Use running means“. Pokud by byla tato možnost zvolena, docházelo by k přepočtu center po přidání jakéhokoliv nového člena. V druhém případě se středy určí až po naplnění shluků všemi objekty. Výsledky metody *k*-průměrů mohou být do značné míry ovlivněny pořadím dat ve vstupním souboru. V práci proto byla využita varianta přepočtu center až po naplnění všemi objekty, která tuto závislost anulovala.



Obrázek 4.6 - Prostředí metody *k* - průměrů v SPSS, zdroj: Vlastní zpracování.

Druhou nedokonalost skrývá metoda ve způsobu volby počátečních center. Proto bylo shlukování opakováno několikrát a z výsledných tří analýz byla vybrána ta varianta, ve které byl součet euklidovské vzdálenosti všech objektů od jim přiřazeným centrům nejmenší. V první variantě bylo stanovení ponecháno na programu SPSS, který jako počáteční centra zahrnoval odlehlejší objekty. V druhé volbě byla zvolena dosavadní centra a další dvě možná řešení a ve třetí byly za počáteční centra stanoveny objekty, které mají k sobě nejblíže. Tyto tabulky jsou rovněž zobrazeny v přiložených excelovských souborech, ve kterých jsou hodnoty zvolených center vyznačeny tučně.

Bohužel přezkoumáním této analýzy nebyly získány jasné výsledky. Pro každé nastavení hodnoty *k* byla zvolena jiná varianta. Není proto vyloučeno, že pokud by byla zvolena jiná počáteční centra, dosahovala by metoda lepších výsledků, než v případě výše stanovených center. Nicméně opakovaná analýza je pro nalezení optimálních výsledků u metody *k*-průměrů doporučována.

Výsledky metody *k*-průměrů

Výsledky aplikace metody *k*-průměrů jsou oproti grafickým možnostem hierarchického shlukování zobrazeny pouze tabulkami. V případě analyzovaných možností jsou rovněž součástí přiložených souborů. Jednotlivé tabulky zobrazují

- zvolená počáteční shlukovací centra,
- změnu shlukovacích center,
- rozdělení objektů do shluků a jejich vzdálenost od centra,
- konečná shlukovací centra,
- vzdálenost mezi výslednými centry
- a ANOVA tabulku.

Tabulka zobrazující počet kroků, ve kterých docházelo ke změnám počátečních center, se napříč variantami liší. V prvním případě bylo nutné centra změnit čtyřikrát, ve druhém sedmkrát a pro stanovených pět shluků bylo těchto kroků šest. Proces končí ve chvílí, kdy změna není provedena u žádného z center. Pokud je tedy kritérium nulové například pouze u jednoho shluku, v dalším kroku je přepočítáváno znovu.

Pokud se zaměříme na konečné rozložení objektů do shluků, lze pro každou zvolenou variantu *k* sledovat nerovnoměrné rozložení objektů. Pro zvolených pět shluků byla dokonce v posledním z nich obsažena pouze tři města. Osamostatnit takovýto malý počet objektů by bylo pro firmu jen těžko realizovatelné. Zejména z tohoto důvodu jsou přílohy dostupné pouze pro tří a čtyř shlukovou variantu.

Oproti hierarchickým metodám jsou v SPSS jako jeden z výstupů uváděna konečná shlukovací centra. Jelikož se vycházelo ze standardizovaných hodnot, bylo nutné tento výstup převést zpět na geografické údaje za pomoci vzorce 2.11. Tabulky 4.11 a 4.12 zobrazují tyto výsledky pro tří a čtyř shlukovou variantu.

Tabulka 4.11 - Geografické hodnoty konečných center pro tří shlukovou variantu

Pro $k = 3$	1. shluk	2. shluk	3. shluk
Zeměpisná šířka	49,8188	50,4830	49,7240
Zeměpisná délka	17,1719	14,6042	13,2972

Tabulka 4.12 - Geografické hodnoty konečných center pro čtyř shlukovou variantu

Pro $k = 4$	1. shluk	2. shluk	3. shluk	4. shluk
Zeměpisná šířka	50,5022	49,9549	49,6981	49,7240
Zeměpisná délka	14,5180	16,2917	18,1022	13,2972

K ověření rozdílnosti mezi shluky byla opět využita analýza rozptylu (tabulka 4.13). Tabulka 4.13 zobrazuje tuto analýzu pro rozložení objektů do tří shluků. Ve srovnání s hierarchickými metodami lze ale F-test využít pouze pro popisné účely, jelikož je vycházeno ze standardizovaných hodnot. Nemůžeme zde tedy testovat hypotézu, zda jsou si průměry shluků rovny.

Tabulka 4.13 - Analýza rozptylu metody *k*-průměrů pro tři shluky

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Zscore(Z_S)	15,065	2	,260	38	57,998	,000
Zscore(Z_D)	14,887	2	,269	38	55,323	,000

4.3.7 Shlukování pomocí reálných vzdáleností

Pokud shrneme výsledky metod z kapitol 4.3.2 až 4.3.6, dostáváme se k závěru, že využití typických měr vzdáleností není pro analýzu zabývající se pouze geografickými daty ideální variantou. Z toho důvodu byla pro další část práce sestrojena matice reálných vzdáleností. Každá hodnota této matice reprezentuje nejrychlejší dopravní vzdálenost mezi dvěma objekty. Data byla čerpána ze serveru mapdevelopers.com pracující na mapách Google.

Naskýtá se zde i možnost využití volby nejkratších vzdáleností. I když v této variantě dochází k částečné úspoře, nastává zde poměrně velký nesoulad s časem, který je pro každý logistický proces rovněž klíčový. Kupříkladu nejkratší cesta z Liberce do Ostravy vede přes Polskou republiku a trvá přes 7 hodin. Pokud je ale zvolena nejrychlejší trasa, prodlouží se tento přesun o pár desítek kilometrů a ušetří se bezmála tři hodiny. K tomuto příkladu se váže i druhý předpoklad a to omezení dopravy pouze na vnitrostátní komunikaci.

Program SPSS dovoluje po menší úpravě syntaxe každé hierarchické metody vložení vlastního souboru, kterým lze nahradit zvolenou metriku. Identifikační čísla z Přílohy č. 1 již neodpovídají přiřazeným městům, jelikož města byla řazena podle získané matice reálných vzdáleností. Ukázka vlastní matice vzdáleností je zobrazena v tabulce 4.14.

Tabulka 4.14 – Ukázka matice reálných vzdáleností (v km)

Case	1:Jeseník	2:Ostrava	3:Opava	4:Valašské Meziříčí
1:Jeseník	,000	112,000	79,200	152,000
2:Ostrava	112,000	,000	32,600	67,400
3:Opava	79,200	32,600	,000	62,500
4:Valašské Meziříčí	152,000	67,400	62,500	,000

Druhou podstatnou změnou při využití těchto metod byla nemožnost sledované hodnoty standardizovat. Jelikož se matice vzdáleností nestanovuje z hodnot proměnných, ale je vložena externě, není nutné standardizaci provádět.

Výběr vhodné varianty

S využitím reálných vzdáleností dosahovaly porovnávané metody podstatně rozdílných výsledků v porovnání s předchozí analýzou, ve které byly vzdálenosti mezi objekty stanoveny na základě euklidovské nebo čtvercové euklidovské metriky. Postupně byly prováděny stejné metody jako v kapitole 4.3. Velice špatných výsledků znovu dosahovala metoda nejbližšího souseda, ve které rovněž docházelo k řetězení a podstatný rozdíl mezi shluky byl patrný až v předposledním kroku aglomerativní tabulky. K obdobným výsledkům, dokonce se snižováním koeficientů podobnosti ve vybraných krocích, docházelo v případě Mediánové metody. Je tedy možné konstatovat, že reálná vzdálenost není pro tuto metodu vhodná.

Srovnatelných výsledků dosahovaly metody Wardova, nevzdálenějšího souseda a metoda průměrné vazby pro vnitroshlukovou vzdálenost (dále jen metoda průměrné vazby). Ve všech třech bylo z koeficientů podobnosti a dendrogramů patrné rozdělení na čtyři shluky. Všechny tři metody měly shodný první shluk. Jsou v něm obsažena města z Moravy a Slezska. Metody Wardova a nejvzdálenějšího souseda se lišily pouze v jednom objektu a to ve městě Kolín. Oproti tomu metoda průměrné vazby byla v posuzovaných třech shlucích rozdílnější. Výsledky těchto metod jsou stejně jako v předchozích případech zobrazeny v excelovských souborech. Výstupy na jednotlivých listech jsou obdobné jako v případě kapitoly 4.3.1, se změnou v matici vzdáleností.

Pro výběr vhodné metody s využitím reálných vzdáleností bylo nutné stanovit konečná skladovací centra pro shluky, ve kterých se tyto metody lišily. Vzhledem k tomu, že hierarchické shlukovací metody s využitím reálných vzdáleností nestanovují v programu SPSS konečná shlukovací centra jako v případě metody *k*-průměrů, byla tato místa pro každý shluk určena pomocí průměrných hodnot dvou využívaných proměnných. Jako jednodušší řešení se naskýtá využití metody *k*-průměrů a stanovit centra pomocí jejího algoritmu. Bohužel tato metoda pracuje v SPSS implicitně s euklidovskou vzdáleností, kterou nebylo možné nahradit.

Dále bylo bráno v úvahu, že nové skladovací prostory budou z hlediska počtu obyvatel umístěny ve větších městech. Důvodem je snazší nalezení vhodného objektu, vyšší počet potencionálních zájemců o práci a v neposlední řadě dobrá dopravní dostupnost. Pro Wardovu metodu a metodu nejvzdálenějšího souseda byla tato centra stanovena ve městech Mělník, Stříbro, Pardubice. Pro metodu průměrné vazby byla za centra zvolena města Litoměřice, Stříbro a Hradec Králové. K porovnání metod byl zvolen prostý součet vzdáleností mezi centrem a jednotlivými městy spojenými s tímto centrem. Z tabulky 4.15 je patrné, že nejkratší vzdálenosti dosahuje třetí z posuzovaných metod, tedy metoda průměrné vazby.

Tabulka 4.15 – Porovnání vzdálenosti (v km)

Metoda	2. shluk	3. shluk	4. shluk	Celkem
Wardova metoda	360,7	960,6	401,9	1723,2
Metoda nejvzdálenějšího souseda	360,7	467,9	855,8	1684,4
Metoda průměrné vazby	452,7	405	696,23	1553,93

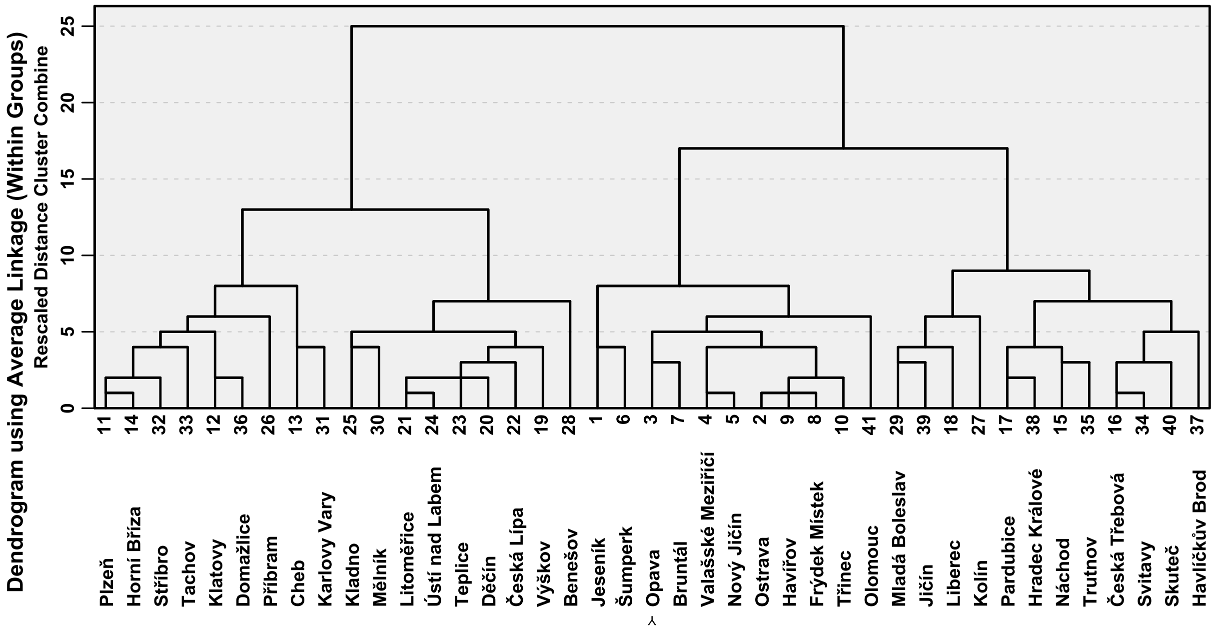
Shrnutí výsledků

Na základě výše stanovených vzdáleností byla za nejvhodnější variantu vybrána metoda průměrné vazby. Pokud se tedy zaměříme na výslednou aglomerativní tabulku (tabulka 4.16), lze pozorovat zvýšení hodnoty koeficientů podobnosti mezi 37. a 38. krokem procesu.

Tabulka 4.16 - Aglomerativní tabulka metody průměrné vazby

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
34	19	28	70,690	29	0	38
35	11	13	77,717	32	18	38
36	1	2	77,995	24	31	39
37	15	18	88,321	33	30	39
38	11	19	123,028	35	34	40
39	1	15	157,799	36	37	40
40	1	11	229,448	39	38	0

Rozdělení objektů do čtyř shluků podporuje dále i výsledný dendrogram (obrázek 4.7) interpretované metody. Pokud by byl na hladině významnosti 10 proveden horizontální řez, vyplývalo by z něj konečné rozdělení do zmiňovaných čtyř shluků. V tabulce 4.17 jsou již jednotlivá města zařazena do shluků podle svého rozdělení. Shluky jsou co do velikosti lépe srovnatelné než u výsledků metod založených na euklidovské, popřípadě čtvercové euklidovské vzdálenosti. I zde však můžeme, zejména ve shlucích jedna a čtyři, sledovat relativně vzdálené objekty. Pro první shluk jsou to stále města Jeseník a Šumperk a ve čtvrtém shluku Benešov.



Obrázek 4.7 - Dendrogram metody průměrné vzdálenosti, zdroj: Vlastní zpracování.

Tabulka 4.17 - Rozdělení objektů dle metody průměrné vzdálenosti

1. shluk (11)	2. shluk (9)	3. shluk (12)	4. shluk (9)
Jeseník	Plzeň	Náchod	Výškov
Ostrava	Klatovy	Česká Třebová	Děčín
Opava	Cheb	Pardubice	Litoměřice
Valašské Meziříčí	Horní Bříza	Liberec	Česká Lípa
Nový Jičín	Příbram	Kolín	Teplice
Šumperk	Karlovy Vary	Mladá Boleslav	Ústí nad Labem
Bruntál	Stříbro	Svitavy	Kladno
Frýdek Místek	Tachov	Trutnov	Benešov
Havířov	Domažlice	Havlíčkův Brod	Mělník
Třinec		Hradec Králové	
Olomouc		Jičín	
		Skuteč	

I v tomto případě lze na základě jednofaktorové analýzy rozptylu (tabulka 4.18) potvrdit statisticky významnou rozdílnost sledovaných shluků.

Tabulka 4.18 – Analýza rozptylu metody průměrné vazby

		Sum of Squares	df	Mean Square	F	Sig.
Z_S	Between Groups	2,936	3	,979	11,071	,000
	Within Groups	3,271	37	,088		
	Total	6,206	40			
Z_D	Between Groups	129,139	3	43,046	181,532	,000
	Within Groups	8,774	37	,237		
	Total	137,912	40			

4.4 Výsledky implementace a zhodnocení výsledků

Při ekonomické interpretaci výsledků této práce je nutné brát zřetel na některá dosavadní zjednodušení oproti realitě. Celková optimalizace logistického procesu není založena pouze na snaze o minimalizaci dopravních vzdáleností. Jsou zde důležité i další parametry a to zejména čas, velikost požadovaných svozů a rozvozů nebo různé organizační a personální důsledky při uplatnění změn. Nicméně i výsledky shlukovací analýzy jsou velkým přínosem a slouží jako jeden z klíčových předpokladů pro odhad budoucího vývoje logistiky modelovaného podniku.

Následné kapitoly tedy nezahrnují v souladu s cílem práce všechny uvedené parametry a jsou soustředěny zejména na nastínění možných variant při využití výsledků shlukovací metody průměrné vazby.

4.4.1 Stanovení nových dopravních tras

Výsledky metody průměrné vazby bylo potřeba doplnit o poslední neuváděné centrum. Vzhledem k velkému rozptylu měst v Moravskoslezském a Olomouckém kraji a rovněž možnosti zásobování města Olomouce při zásobování Ostravy, bylo zachováno v případě prvního shluku původní centrum. Tabulka 4.19 zobrazuje zvolená centra spolu s průměrnými hodnotami geografických údajů pro dané shluky. Dále zobrazuje euklidovskou vzdálenost mezi průměrnou hodnotou a zvoleným centrem.

Tabulka 4.19 – Shlukovací centra vycházející z metody průměrné vazby

Shluk	Počet objektů	Zeměpisná šířka	Zeměpisná délka	Centrum	Euklid. vzdálenost (km)
1	11	49,7971	17,8598	Ostrava	29,5
2	9	49,7730	13,0935	Stříbro	7,1
3	12	50,1644	15,7293	Hradec Králové	7,7
4	9	50,4427	14,1791	Litoměřice	10,9

Prvořadě pro práci nebylo nutné, aby nová skladovací centra byla vytyčena v místech center starých, popřípadě služeben. Na druhou stranu je ale tato varianta vhodnější jednak z hlediska vzdálenosti, tak i rozvozů. Odpadnou totiž závozy do míst, ve kterých se centra nachází. Ostatně v původní variantě tomu nebylo jinak. V nové variantě ale přibyde trasa pro zásobování center z centrálního skladu, který byl z hlediska největšího počtu objektů a vhodné geografické polohy stanoven pro centrum třetího shluku.

Pro výpočet nových tras byla dále brána v potaz maximální možná ujetá vzdálenost (z důvodu běžné pracovní doby a průměrného času potřebného pro nakládku a vykládku) a průměrný počet požadavků (v tabulkách uváděno jako náklad) pro každou služebnu. Tyto hodnoty jsou stanoveny za období duben 2015 až únor 2016. Pro maximální ujetou vzdálenost v rámci jedné trasy byla tedy zvolena hodnota 320 km. Počet požadavků je zobrazen u každé z variant. Takto zobecněný problém byl řešen pomocí SW dostupného na serveru www.logvrp.com, který nastínil nové přepravní možnosti. Výsledné tabulky obsahující stanovené trasy, počty kontejnerů k přepravě a vzdálenosti mezi městy jsou zobrazeny v Příloze č. 2.

4.4.2 Ekonomické zhodnocení modelu

Konečné zhodnocení výsledků je založeno na porovnání celkového součtu vzdáleností logistického procesu pro původní a novou variantu. Jak již bylo zmíněno, komplexní posouzení problému přesahuje svou rozsáhlostí možnosti této práce. Na druhou stranu jsem toho názoru, že i pouhá úspora ve vzdálenosti (tím pádem i úspora nákladů na dopravu) je pro firmu jeden z klíčových ukazatelů pro posouzení dalšího vývoje. V případě součtu vzdáleností pro původní stav bylo nutné všechny trasy uvedené v kapitole 3.2 nahradit hodnotami z matice reálných vzdáleností. Délky jízd uváděné v popisu původní varianty obsahují totiž maximální možnou, nebo v případě interní dopravy, průměrnou vzdálenost každé z nich. Druhý sloupec tabulky 4.20 zobrazuje pro každé centrum součet všech jízd, čtvrtý sloupec doplňuje tyto závozy o komunikaci mezi centrálním skladem a zbylými dvěma centry.

Tabulka 4.20 – Součet původních vzdáleností (v km)

Centrum	Celková vzdálenost	Trasy mezi centry	Celková vzdálenost
Skuteč	4000,2	Skuteč - Jičín - Liberec	295,6
Liberec	713,2	Skuteč - Liberec	368
Ostrava	550,5	Skuteč - Ostrava	428
		Skuteč - Olomouc - Ostrava	420,5
Celkem			**6776**

Obdobně lze popsat i tabulku 4.21 pro variantu založenou na metodě průměrné vazby. Z tabulky je patrné, že se výrazně snížila vzdálenost pro obsluhu služeben z nově vzniklých center, naopak vzrostla vzdálenost mezi centrálním skladem a centry ostatních shluků.

Tabulka 4.21 – Součet vzdáleností optimalizované varianty (v km)

Centrum	Celková vzdálenost	Trasy mezi centry	Celková vzdálenost
Hradec Králové	809,5	Hradec Králové - Litoměřice	680
Litoměřice	620	Hradec Králové - Stříbro	998
Ostrava	550,5	Hradec Králové - Ostrava	478
Stříbro	546,3	Hradec Králové - Olomouc - Ostrava	472,8
Celkem			**5155,1**

Z posledních řádků obou výše uváděných tabulek je patrné, že v případě využití výsledků metody průměrné vazby a dopravních tras nastíněných v Příloze č. 2, dochází v porovnání s původní variantou k měsíční úspoře o bezmála 1621 km. Celková vzdálenost může do jisté míry mírně zkreslovat, jelikož trasy pro rozvoz a následný svoz nemusí být totožné. Z toho vyplývá, že tento souhrn by měl být brán jen jako odhad úspor.

5 Závěr

Cílem této práce bylo porovnat dosavadní stav logistického procesu ve vybrané firmě s řešením, které by optimalizovalo její logistický tok. Jako prostředek pro splnění cíle této práce byly využity shlukovací algoritmy.

V teoretické části byly vysvětleny základní principy týkající se shlukování, způsobu měření vzdálenosti mezi shluky spolu s popisem dat, pro která jsou tyto míry stanoveny. V další části byly nastíněny principy hierarchických a nehierarchických metod shlukování a porovnány jejich zásadní rozdíly. Rovněž byly uvedeny i postupy pro rozsáhlé datové soubory a nové shlukovací algoritmy.

Po provedení popisu současného stavu v třetí části práce, který byl potřebný k porozumění problematiky, následovala analýza s využitím softwaru IBM SPSS Statistics. V práci byly využity hierarchické metody využívající euklidovskou, popřípadě čtvercovou euklidovskou vzdálenost a metoda *k*-průměrů. Dále byl vysvětlen celkový průběh shlukování a detailně rozebrány výstupy analyzovaných metod. Pro přiblížení se reálnému stavu byla sestrojena matice reálných vzdáleností, ze které bylo shlukování za pomoci vybraných hierarchických metod opakováno. Výsledné rozvržení objektů do shluků se v porovnání s výše uváděnými mírami vzdálenosti podstatně lišilo. Nejlepších výsledků dosahovala metoda průměrné vazby pro vnitroshlukovou vzdálenost. Výsledné shluky obsahovaly srovnatelný počet objektů a v porovnání s ostatními metodami byl součet vzdáleností od nově vzniklých center nejmenší.

V poslední částí byly implementovány výsledky metody průměrné vazby pro vnitroshlukovou vzdálenost. Za nová centra byla stanovena města Hradec Králové, Stříbro, Ostrava a Litoměřice. Dále byly nastíněny nové závozní trasy a ty byly porovnány s dosavadním stavem. Výsledkem práce je měsíční odhad úspor v dopravě o bezmála 1621 km.

Z hlediska rozsáhlosti úlohy nebylo možné do práce zahrnout veškeré aspekty pro stanovení nových logistických toků v modelovaném podniku. Nicméně struktura práce odpovídá definovanému cíli, byly zde stanoveny nové centrální sklady a rozvozové trasy. Na základě tohoto schématu mohou být v budoucnu využity ostatní parametry pro úlohy typu dopravních problémů nebo při tvorbě komplexních algoritmů.

Seznam použité literatury

Literatura:

BUDÍKOVÁ, M., T. LERCH a Š. MIKOLÁŠ. *Základní statistické metody*. Brno: Masarykova Univerzita, 2005. ISBN 80-210-3886-1.

HAN, J., M. KAMBER and J. PEI. *Data Mining: Concepts and Techniques*. 2. ed. San Francisco: Morgan Kaufmann Publishers Inc, 2006. ISBN 1-55860-901-6.

KOŠŤÁL, Jaroslav. *Vybrané metody vícerozměrné statistiky*. Praha: Institut pro kriminologii a sociální prevenci, 2013. ISBN 978-80-7338-128-8. Dostupné také z http://lpxp.sweb.cz/403.pdf

LUKASOVÁ, Alena a Jana ŠARMANOVÁ. *Metody shlukové analýzy*. Praha: SNTL, 1985. ISBN 04-014-85.

MELOUN Milan a Jiří MILITKÝ. *Statická analýza experimentálních dat*. Praha: ACADEMIA, 2004. ISBN 80-200-1254-0.

PETR, Pavel. *Stručný návod k ovládání IBM SPSS Statistics a IBM SPSS Modeler*. Pardubice: Univerzita Pardubice Fakulta ekonomicko-správní, 2012. ISBN 978-80-7395-477-2. Dostupné také z: http://homel.vsb.cz/~kur138/ADDS/PetrP_IBM_Statistics_2012.pdf

ŘEZANKOVÁ, H., D. HÚSEK a V. SNÁŠEL. *Shluková analýza dat*. Praha: Professional Publishing, 2007. ISBN 978-80-86946-26-9.

Odborné články:

RYBOVÁ, Jarmila. Shluková analýza v problematice daní. In: *Acta oeconomica Pragensia* [online]. 2015, roč. 23, č. 3, s. 58-66 [cit. 2016-1-15]. ISSN 0572-3043. Dostupné z: https://www.vse.cz/aop/476

Elektronické zdroje:

ČEZ, a. s. O společnosti. *CEZ.cz* [online]. ©2016 [cit. 2016-03-02]. Dostupné z: https://www.cez.cz/cs/o-spolecnosti/cez/profil-spolecnosti.html

ČEZ Distribuční služby, s. r. o. Profil společnosti. *CEZ.cz* [online]. ©2016 [cit. 2016-03-02]. Dostupné z: https://www.cez.cz/cds/cs/o-spolecnosti/profil-spolecnosti.html

HORÁK, Jiří. *Prostorová analýza dat* [online]. 2002 [cit. 2015-12-20]. Dostupné z http://gis.vsb.cz/pad/index.htm

HYNAR, Martin. *Metody shlukování* [online]. 2003 [cit. 2016-10-02]. Dostupné z: http://www.fit.vutbr.cz/study/courses/ZZD/public/seminar0304/Shlukovani1-text.pdf

JARKOVSKÝ, Jiří a Simona LITTNEROVÁ. *Shluková analýza* [online prezentace]. Brno: Masarykova univerzita v Brně. [cit. 2015-12-22]. Dostupné z: http://www.iba.muni.cz/esf/index.php?pg=bimat--predmety-bakalarskeho-studia--vicerozmerne-statisticke-metody

KELBEL, Jan a David ŠILHÁN. *Shluková analýza* [online]. [cit. 2015-12-22]. Dostupné z: http://cmp.felk.cvut.cz/cmp/courses/recognition/zapis_prednasky/zapis_02/13/shlukovani.pdf

Logvrp – The Route Planner [online]. Logvrp.com ©2012. [cit. 2016-3-27]. Dostupné z: http://logvrp.com/logvrpsite/Default.aspx

MARŠÁLKOVÁ, Kateřina. *Hodnocení úspěšnosti metod shlukové analýzy.* Praha, 2015. Diplomová práce. Vysoká škola ekonomická v Praze. Dostupné z: https://www.vse.cz/vskp/46551_hodnoceni_uspesnosti_metod_shlukove_analyzy

PIECH, Chris. *K Means* [online]. 2013 [cit. 2015-12-11]. Dostupné z: http://stanford.edu/~cpiech/cs221/handouts/kmeans.html

ZAIANE, Osmar. *Chapter8: Data Clustering* [online]. 1999 [cit. 2016-01-20]. Dostupné z: https://webdocs.cs.ualberta.ca/~zaiane/courses/cmput690/slides/Chapter8/index2.html

ZHANG, T., R. RAMAKRISHNAN and M. LIVNY. BIRCH: an efficient data clustering method for very large databases. *ACM Sigmod Record* [online]. 1996 [cit. 2016-1-15]. Dostupné z: http://www.cs.sfu.ca/CourseCentral/459/han/papers/zhang96.pdf

ŽAMBOCHOVÁ, Marta. *Shluková analýza rozsáhlých souborů dat: nové postupy založené na metodě k-průměrů*. Praha, 2010. Disertační práce. Vysoká škola ekonomická v Praze, Fakulta informatiky a statistiky. Dostupné z: https://www.vse.cz/vskp/eid/27036

Seznam použitých zkratek

AMS	Autorizované metrologické středisko a opravna
BIRCH	Balanced Iterative Reducing and Clustering Using Hierarchies
CF	Clustering Feature
CLARA	Clustering Large Application
CLARANS	Clustering Large Application based on Randomised Search
CURE	Clustering Using Representatives
ČDS	ČEZ Distribuční služby
ISODATA	Iterative Self-Organizing Data Analysis Techniques
ODBC	Open Database Connectivity
PAM	Partitioning Around Medoids
SLM	Správa a logistika měřidel
SW	Software

Prohlášení o využití výsledků bakalářské práce

Prohlašuji, že

- jsem byl seznámen s tím, že na mou bakalářskou práci se plně vztahuje zákon č. 121/2000 Sb. – autorský zákon, zejména § 35 – užití díla v rámci občanských a náboženských obřadů, v rámci školních představení a užití díla školního a § 60 – školní dílo;

- beru na vědomí, že Vysoká škola báňská – Technická univerzita Ostrava (dále jen VŠB-TUO) má právo nevýdělečně, ke své vnitřní potřebě, bakalářskou práci užít (§ 35 odst. 3);

- souhlasím s tím, že bakalářská práce bude v elektronické podobě archivována v Ústřední knihovně VŠB-TUO a jeden výtisk bude uložen u vedoucího bakalářské práce. Souhlasím s tím, že bibliografické údaje o bakalářské práci budou zveřejněny v informačním systému VŠB-TUO;

- bylo sjednáno, že s VŠB-TUO, v případě zájmu z její strany, uzavřu licenční smlouvu s oprávněním užít dílo v rozsahu § 12 odst. 4 autorského zákona;

- bylo sjednáno, že užít své dílo, bakalářskou práci, nebo poskytnout licenci k jejímu využití mohu jen se souhlasem VŠB-TUO, která je oprávněna v takovém případě ode mne požadovat přiměřený příspěvek na úhradu nákladů, které byly VŠB-TUO na vytvoření díla vynaloženy (až do jejich skutečné výše).

V Ostravě dne 6. 5. 2016

Jan Timel

Jan Timel

Seznam příloh

Příloha č. 1: Tabulka vstupních dat

Příloha č. 2: Součet vzdáleností nových dopravních tras

Přílohy

Příloha č. 1: Tabulka vstupních dat

ID	Město	Zeměpisná šířka	Zeměpisná délka
1	Česká Lípa	50,6751153	14,5205081
2	Děčín	50,7756572	14,1999742
3	Liberec	50,7634033	15,050605
4	Litoměřice	50,5449411	14,1125414
5	Výškov	50,3954228	13,6720992
6	Teplice	50,6425075	13,8172797
7	Ústí nad Labem	50,6608403	14,0320806
8	Bruntál	49,9846256	17,4626331
9	Frýdek Místek	49,6810306	18,3441547
10	Havířov	49,7993561	18,4099456
11	Nový Jičín	49,5996839	18,0089219
12	Olomouc	49,597185	17,2547536
13	Opava	49,9417497	17,8959575
14	Ostrava	49,8277811	18,2651347
15	Šumperk	49,9669831	16,9714694
16	Jeseník	50,2314822	17,2060264
17	Třinec	49,66809	18,6718425
18	Valašské Meziříčí	49,4706758	17,9669961
19	Benešov	49,7884425	14,6821067
20	Kladno	50,1431489	14,1086769
21	Kolín	50,0230272	15,2119839
22	Mělník	50,363805	14,4731447
23	Mladá Boleslav	50,4124975	14,9334519
24	Příbram	49,6756581	13,9951989
25	Česká Třebová	49,8922414	16,4517325
26	Havlíčkův Brod	49,6101014	15,5560483
27	Hradec Králové	50,2152147	15,8122028
28	Jičín	50,4352608	15,3604442
29	Náchod	50,4187119	16,1707514
30	Pardubice	50,0333692	15,7779511
31	Svitavy	49,7528408	16,4730042
32	Trutnov	50,5764772	15,9583764
33	Domažlice	49,4387317	12,9350606
34	Horní Bříza	49,8438694	13,3637625
35	Cheb	50,0828458	12,3779092
36	Karlovy Vary	50,2286394	12,8487964
37	Klatovy	49,3993294	13,2900406
38	Plzeň	49,7423478	13,3940497
39	Stříbro	49,7529375	13,0040917
40	Tachov	49,7924314	12,6328214
41	Skuteč	49,8395236	15,9947306

Příloha č. 2: Součet vzdáleností nových dopravních tras

Centrum	Závoz	Služebny	Náklad (ks)	Vzdálenost (km)
Ostrava	1.	Ostrava		
		Nový Jičín	3	45,7
		Valašské Meziříčí	4	18,3
		Ostrava		67,4
		Celkem	**7**	**131,4**
	2.	Ostrava		
		Opava	3	32,6
		Bruntál	2	36,8
		Jeseník	1	43,3
		Šumperk	2	47,8
		Ostrava		157
		Celkem	**8**	**317,5**
	3.	Ostrava		
		Havířov	3	17,2
		Třinec	3	31,5
		Frýdek Místek	4	32,5
		Ostrava		20,4
		Celkem	**10**	**101,6**

Centrum	Závoz	Služebny	Náklad (ks)	Vzdálenost (km)
Stříbro	1.	Stříbro		
		Horní Bříza	3	37,5
		Karlovy Vary	4	78
		Cheb	4	41,7
		Tachov	2	49
		Stříbro		30,7
		Celkem	**13**	**236,9**
	2.	Stříbro		
		Příbram	5	92
		Klatovy	5	70,3
		Domažlice	2	30,6
		Stříbro		41,5
		Celkem	**12**	**234,4**
	3.	Stříbro		
		Plzeň	6	37,5
		Stříbro		37,5
		Celkem	**6**	**75**

Centrum	Závoz	Služebny	Náklad (ks)	Vzdálenost (km)
Litoměřice	1.	Litoměřice		
		Teplice	6	36,8
		Výškov	10	41
		Litoměřice		43,4
		Celkem	**16**	**121,2**
	2.	Litoměřice		
		Ústí nad Labem	5	18,6
		Děčín	5	24
		Česká Lípa	5	33
		Litoměřice		38,1
		Celkem	**15**	**113,7**
	3.	Litoměřice		
		Kladno	11	65
		Litoměřice		65
		Celkem	**11**	**130**
	4.	Litoměřice		
		Mělník	5	50
		Benešov	8	91,6
		Litoměřice		113,5
		Celkem	**13**	**255,1**

Centrum	Závoz	Služebny	Náklad (ks)	Vzdálenost (km)
Hradec Králové	1.	Hradec Králové		
		Pardubice	4	29
		Skuteč	4	33
		Svitavy	2	47,8
		Česká Třebová	3	21
		Hradec Králové		64
		Celkem	**13**	**194,8**
	2.	Hradec Králové		
		Náchod	4	42,4
		Trutnov	2	35
		Jičín	3	50,7
		Hradec Králové		45
		Celkem	**9**	**173,6**
	3.	Hradec Králové		
		Mladá Boleslav	4	80,4
		Liberec	10	50
		Hradec Králové		100
		Celkem	**14**	**230,4**
	4.	Hradec Králové		
		Kolín	10	60,7
		Havlíčkův Brod	4	58
		Hradec Králové		92,5
		Celkem	**14**	**211,2**